J Neurophysiol 120: 2555–2570, 2018.
First published August 29, 2018; doi:10.1152/jn.00222.2018.

RESEARCH ARTICLE | *Higher Neural Functions and Behavior*

High-level language processing regions are not engaged in action observation or imitation

Brianna L. Pritchett,[1] Caitlyn Hoeflin,[1] Kami Koldewyn,[2] Eyal Dechter,[1] and Evelina Fedorenko[1,3,4]
[1]*Department of Brain and Cognitive Sciences/McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts;* [2]*School of Psychology, Bangor University, Gwynedd, United Kingdom;* [3]*Department of Psychiatry, Massachusetts General Hospital, Charlestown, Massachusetts; and* [4]*Department of Psychiatry, Harvard Medical School, Boston, Massachusetts*

Submitted 2 April 2018; accepted in final form 22 August 2018

Pritchett BL, Hoeflin C, Koldewyn K, Dechter E, Fedorenko E. High-level language processing regions are not engaged in action observation or imitation. *J Neurophysiol* 120: 2555–2570, 2018. First published August 29, 2018; doi:10.1152/jn.00222.2018.—A set of left frontal, temporal, and parietal brain regions respond robustly during language comprehension and production (e.g., Fedorenko E, Hsieh PJ, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N. *J Neurophysiol* 104: 1177–1194, 2010; Menenti L, Gierhan SM, Segaert K, Hagoort P. *Psychol Sci* 22: 1173–1182, 2011). These regions have been further shown to be selective for language relative to other cognitive processes, including arithmetic, aspects of executive function, and music perception (e.g., Fedorenko E, Behr MK, Kanwisher N. *Proc Natl Acad Sci USA* 108: 16428–16433, 2011; Monti MM, Osherson DN. *Brain Res* 1428: 33–42, 2012). However, one claim about overlap between language and nonlinguistic cognition remains prominent. In particular, some have argued that language processing shares computational demands with action observation and/or execution (e.g., Rizzolatti G, Arbib MA. *Trends Neurosci* 21: 188–194, 1998; Koechlin E, Jubault T. *Neuron* 50: 963–974, 2006; Tettamanti M, Weniger D. *Cortex* 42: 491–494, 2006). However, the evidence for these claims is indirect, based on observing activation for language and action tasks within the same broad anatomical areas (e.g., on the lateral surface of the left frontal lobe). To test whether language indeed shares machinery with action observation/execution, we examined the responses of language brain regions, defined functionally in each individual participant (Fedorenko E, Hsieh PJ, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N. *J Neurophysiol* 104: 1177–1194, 2010) to action observation (*experiments 1*, *2*, and *3a*) and action imitation (*experiment 3b*). With the exception of the language region in the angular gyrus, all language regions, including those in the inferior frontal gyrus (within "Broca's area"), showed little or no response during action observation/imitation. These results add to the growing body of literature suggesting that high-level language regions are highly selective for language processing (see Fedorenko E, Varley R. *Ann NY Acad Sci* 1369: 132–153, 2016 for a review).

NEW & NOTEWORTHY Many have argued for overlap in the machinery used to interpret language and others' actions, either because action observation was a precursor to linguistic communication or because both require interpreting hierarchically-structured stimuli. However, existing evidence is indirect, relying on group analyses or reverse inference. We examined responses to action observation in language regions defined functionally in individual participants and found no response. Thus language comprehension and action observation recruit distinct circuits in the modern brain.

action imitation; action observation; fMRI; functional specificity; language network

INTRODUCTION

Although brain regions that support high-level language processing have been shown to be selective for language over various nonlinguistic cognitive processes (e.g., Fedorenko and Varley 2016), the idea of overlap between language processing and action observation and/or execution remains prominent in the literature. Two lines of theorizing have been used to argue for this overlap. The first stemmed from the discovery of mirror neurons in the prefrontal cortex of rhesus macaques. These neurons fire both when a monkey performs an action and when it observes the action performed (Rizzolatti et al. 1988). Rizzolatti and Arbib (1998) and Arbib (2005, 2010) (see also Corballis 2010 and Petrides and Pandya 2009) speculated that in our primate ancestors mirror neurons were critical for understanding one another's actions: a core component of social cognition. They argued that, over time, basic manual actions grew more abstract and eventually became signs, which, in turn, became mouth movements/vocalizations. Thus manual actions are argued to be a fundamental precursor to linguistic communication, and action understanding and language comprehension should share a common neural substrate because they share a common functional ancestor.

Although the general idea that language arose from gesture finds substantial support (e.g., Corballis 2003; Tomasello 2008; cf. Slocombe 2015), the role of mirror neurons in the evolution of language remains debated (e.g., Hickok 2009). The existence of brain cells/regions with properties of the macaque mirror neuron system in humans is supported by a number of studies (e.g., Mukamel et al. 2010; see Molenberghs et al. 2012 for a meta-analysis) but has not gone unchallenged (e.g., Dinstein et al. 2007; Lingnau et al. 2009). Regardless of these controversies, however, given the prominence of the gesture-based hypothesis of language evolution, it seems important to test whether any parts of the language network in the modern human brain respond to action observation/execution.

Address for reprint requests and other correspondence: E. Fedorenko, 43 Vassar St., 46-3037F, Cambridge MA 02139 (e-mail: evelina9@mit.edu).

The second general line of reasoning is that both the language system and the action observation system (possibly restricted to biological actions; e.g., Clerget et al. 2009; Fazio et al. 2009) rely on an amodal mechanism that recognizes and produces hierarchical structure (e.g., Fiebach and Schubotz 2006; Koechlin and Jubault 2006; Tettamanti and Weniger 2006). This mechanism has been argued to reside in the inferior frontal gyrus (IFG), in or around "Broca's area" (we use quotations because the definition of this brain region in the literature is extremely variable and the term has been argued by some to no longer be meaningful as a result; Tremblay and Dick 2016). However, the evidence for overlap between language and action observation in the IFG is problematic because the IFG is among the most structurally (e.g., Amunts et al. 2010) and functionally (e.g., Fedorenko et al. 2012a) heterogeneous brain regions. Furthermore, lateral frontal lobes are characterized by high interindividual variability (e.g., Amunts et al. 1999; Juch et al. 2005; Tomaiuolo et al. 1999). Thus activation overlap between language and action observation in a traditional functional (f)MRI group analysis (e.g., Higuchi et al. 2009), where activations are averaged across individuals, can be misleading (e.g., Nieto-Castañón and Fedorenko 2012), particularly in the aforementioned regions.

Furthermore, some prior studies did not even include a direct within-experiment comparison between a language and an action task (e.g., Binkofski et al. 2000; Clerget et al. 2009; Meister and Iacaboni 2007) and relied solely on the fallacious reverse inference (Poldrack 2006, 2011) to interpret the frontal activations for action tasks. This approach is especially problematic in this case because frontal lobes, including "Broca's area" itself (Fedorenko et al. 2012a), contain both *1*) language-selective regions, and *2*) highly domain-general ones that belong to the fronto-parietal multiple demand (MD) network (e.g., Duncan 2010) and are driven by diverse cognitive demands (e.g., Duncan and Owen 2000; Fedorenko et al. 2013). Thus interpreting frontal activations for an action observation task as reflecting the recruitment of the language system is not justified. Similarly, although many aphasic patients with frontal lesions exhibit deficits in action observation/execution (e.g., Kimura 1977; Kimura et al. 1976; Papagno et al. 1993; Saygin et al. 2004), these patients' lesions are often extensive and plausibly affect two or more functionally distinct regions (cf. Sirigu et al. 1998). Thus arguing for overlap in mechanisms that support language processing and action observation based on such data is also not warranted.

To test, in the most direct way, whether action observation/execution relies on some of the same neural mechanisms as high-level language processing, we examined responses to action observation and imitation in the language regions functionally defined in each individual. This analytic approach circumvents the problem of high interindividual variability in the precise locations of functional regions (e.g., Fischl et al. 2008; Frost and Goebel 2012; Tahmasebi et al. 2012) and thus stands a chance to conclusively answer the question about whether language regions support some aspects of action observation. It is worth noting that this question is conceptually distinct from the question that is at the core of the embodiment debate (see Leshinskaya and Caramazza 2016 for a recent review): namely, whether concepts are "grounded" in sensory-motor systems. We elaborate further on the relationship between these questions in the DISCUSSION.

MATERIALS AND METHODS

The general approach adopted here across the four experiments is as follows: first, we identify the language network in each participant individually using a functional localizer task based on a broad contrast between the reading of sentences vs. sequences of nonwords (Fedorenko et al. 2010). Then, we examine the engagement of these language-responsive voxels in action observation/imitation across several paradigms. This approach has been previously shown to yield higher sensitivity and functional resolution than traditional group-based analyses, as well as more accurate estimates of effect sizes (e.g., Nieto-Castañón and Fedorenko 2012; Saxe et al. 2006). Furthermore, this approach makes the results directly comparable across the four experiments.

It is worth emphasizing that we here focus on high-level language processing regions, i.e., brain regions that support lexico-semantic and combinatorial (semantic and syntactic) processing (e.g., Bautista and Wilson 2016; Blank et al. 2016; Fedorenko et al. 2012b). These regions plausibly underlie our ability to infer meanings from others' linguistic utterances during comprehension as well as to convert our thoughts into linguistic forms during production. This high-level language network is distinct from both lower level perceptual regions that respond selectively to speech, but are not sensitive to the meaningfulness of the speech signal (e.g., Norman-Haignere et al. 2015; Overath et al. 2015) and lower level speech articulation regions that respond robustly when we produce speech sounds but again are not sensitive to the meaningfulness of the utterance (e.g., Basilakos et al. 2017; Bohland and Guenther 2006; Flinker et al. 2015). Thus our main conclusions pertain to the high-level component of the extended language network. We return to this issue in RESULTS.

Participants

Participants were recruited from MIT and the surrounding Cambridge/Boston, MA community and were paid for their participation. Eleven participants were tested in *experiment 1*, 57 in *experiment 2*, 13 in *experiment 3a*, and 16 in *experiment 3b*. Seven participants were excluded (3 for excessive motion, all in *experiment 3b*, 2 for equipment failure, 1 because an incorrect scanner sequence was used, and 1 due to experimenter error), leaving 90 participants for analysis (10 in *experiment 1*, 54 in *experiment 2*, 13 in *experiment 3a*, and 13 in *experiment 3b*). (The number of participants in *experiment 2* was so large because this experiment was used across multiple projects, and we decided to include here all the data available.) Due to some overlap in participants across experiments (8 participated in both *experiment 2* and *3a*, and 5 participated in both *experiment 2* and *3b*), there were 77 unique individuals (age 18–52, mean age 24, 43 females), 68 right-handed (as determined by the Edinburgh handedness inventory, Oldfield 1971, for $n = 69$, or self-report). No participants were excluded based on handedness because we would like to generalize our results to the entire population, as opposed to only the right-handed participants (see Willems et al. 2014 for discussion). The nine left-handed participants all had a left-lateralized language network, as determined by the language localizer task described below. To determine lateralization, the number of language-contrast-activated voxels in the right hemisphere at a fixed significance threshold was subtracted from the number of language voxels in the left hemisphere at the same threshold, and the resulting value was divided by the sum of language voxels across hemispheres (see Mahowald and Fedorenko 2016 for further details on this method). All were native speakers of English, had normal hearing and vision, and had no history of language impairment. The protocol for these studies was submitted to, and approved by, MIT's Committee on the Use of Humans as Experimental Subjects. All participants gave written informed consent in accordance with the requirements of this protocol.

Design and Procedure Common to All Four Experiments

Each participant completed a language localizer task (Fedorenko et al. 2010) and an action observation/imitation task. Twelve participants completed the localizer task in a separate scanning session; the remaining 78 participants performed the localizer and an action experiment in the same session, along with one or two additional tasks for unrelated studies. The entire scanning session lasted for ~2 h. The task used to localize the language network is described in detail in Fedorenko et al. (2010); the materials and scripts are available from the Fedorenko Laboratory website (https://evlab.mit.edu/funcloc). Briefly, we used a reading task contrasting sentences (e.g., THE SPEECH THAT THE POLITICIAN PREPARED WAS TOO LONG FOR THE MEETING) and lists of unconnected, pronounceable nonwords (e.g., LAS TUPING CUSARISTS FICK PRELL PRONT CRE POME VILLPA OLP WORNETIST CHO) in a standard blocked design with a counterbalanced order across runs (for timing parameters, see Table 1). The sentences > nonwords contrast targets brain regions that support lexico-semantic and combinatorial (semantic and syntactic) processing. Stimuli were presented one word/nonword at a time. For 10 participants (in *experiment 1*), each trial ended with a memory probe and they had to indicate, via a button press, whether or not that probe had appeared in the preceding sequence of words/nonwords. The remaining participants instead read the materials passively (we included a button-pressing task at the end of each trial, to help participants remain alert). Importantly, this localizer has been shown to generalize across task manipulations: the sentences > nonwords contrast, and similar contrasts between language and a linguistically degraded control condition, robustly activate the frontotemporal language network regardless of the task, materials, and modality of presentation (e.g., Fedorenko et al. 2010; Fedorenko 2014; Scott et al. 2017).

The action observation tasks included a variety of conditions, including hand actions with (*experiment 1*) or without (*experiment 3a*) a manipulable object, actions that involve different body parts including hands, but also arms, legs, feet, torso, and head (*experiment 2*), face actions (*experiments 2* and *3a*), and specifically eye and mouth actions (*experiment 3a*); the action imitation task similarly included several conditions (*experiment 3b*). We describe each experiment in more detail below.

Table 1. *Timing parameters for the different versions of the language localizer task*

	Version		
	A	B	C
Number of participants	5	5	80
Task: passive reading (PR) or memory (M)?	M	M	PR
Words/nonwords per trial	8	12	12
Trial duration, ms	4,800	6,000	6,000
Fixation	300	300	100
Presentation of each word/nonword	350	350	450
Fixation			500
Memory probe	1,350	1,000	
Fixation	350	500	
Trials per block	5	3	3
Block duration, s	24	18	18
Blocks per condition (per run)	8	8	8
Fixation block duration, s	16	18	14
Number of fixation blocks	5	5	5
Total run time, s	464	378	358
Number of runs	2	2	2

Experiment 1: Hand Action Observation

Participants watched short videos where a small nonnameable three-dimensional object was manipulated in some way by a hand, in a blocked design, and performed a simple one-back task designed to draw attention to the action or the object. (We used nonnameable objects to avoid a potential confound of activating the names of common objects, which would likely elicit some response in the language regions but not due to overlap in computational demands between language understanding and action observation.) In the action condition, participants had to press a button when they saw the same action twice in a row, and in the object condition, they watched the same videos but had to press a button when they saw the same object twice in a row. The task manipulation was included in an effort to maximally focus the participants' attention on the actions in the action condition.

Materials. There were 8 possible hand actions (e.g., push forward with back of the fingers, or pick up with an index finger and a thumb) and 8 possible nonnameable objects, resulting in 64 unique stimuli (see Fig. 1 for screenshots from sample stimuli). A short video was created for each action/object combination. Each video started with the object sitting on a table, and then the hand entered the frame (always from the same side), performed the action, and exited the frame. Because objects take less time to identify than actions (given that actions unfold over time), some steps were taken to make the conditions comparable in difficulty. First, the videos were edited so that the action started as quickly as possible after the onset of the video (on average, the action took ~250 ms to initiate). Second, objects were grouped into "families" for presentation purposes such that objects within a family were visually similar to one another. Conversely, actions were grouped in a way such that actions within a set were visually dissimilar.

Procedure. Each video (trial) lasted 3 s, and trials were grouped into blocks of eight trials each. Each block was preceded by a 2-s instructions screen telling participants which condition they were about to see. Each run consisted of 16 such experimental blocks (26 s each; 8 blocks per condition) and 5 fixation blocks (16 s each, placed at the beginning of the run, and after each set of 4 blocks). Each run thus lasted 496 s (8 min 16 s). Each participant saw either four or five runs. The order of conditions was counterbalanced across runs and participants.

Experiment 2: Face and Body Action Observation

Participants passively watched silent videos of *1*) face actions, *2*) body actions, *3*) driving through natural scenes, *4*) moving man-made objects, and *5*) spatially scrambled versions of these objects in a blocked design (see Pitcher et al. 2011, for a detailed description). For the purposes of the current study, we examined the first two conditions: face actions and body actions. Participants were instructed to watch attentively.

Materials. There were 60 unique stimuli per condition. The videos depicted children moving against a black background. These children performed a variety of actions like dancing, walking, and crawling (see Fig. 1 for screenshots from sample stimuli). The face action videos featured a child's face in motion: smiling, laughing, talking, or looking at someone off-camera. The body action videos featured a child's moving body part (hands, arms, legs, feet, torso, or back of the head) but did not include the face.

Procedure. Each trial consisted of a single video that lasted 3 s, and trials were grouped into blocks of six trials each. Each run consisted of 10 experimental blocks (18 s each; 2 blocks per condition) and 3 fixation blocks (18 s each), placed at the beginning, middle, and end of the run. Each run thus lasted 234 s (3 min 54 s). Each participant saw between 2 and 4 runs.

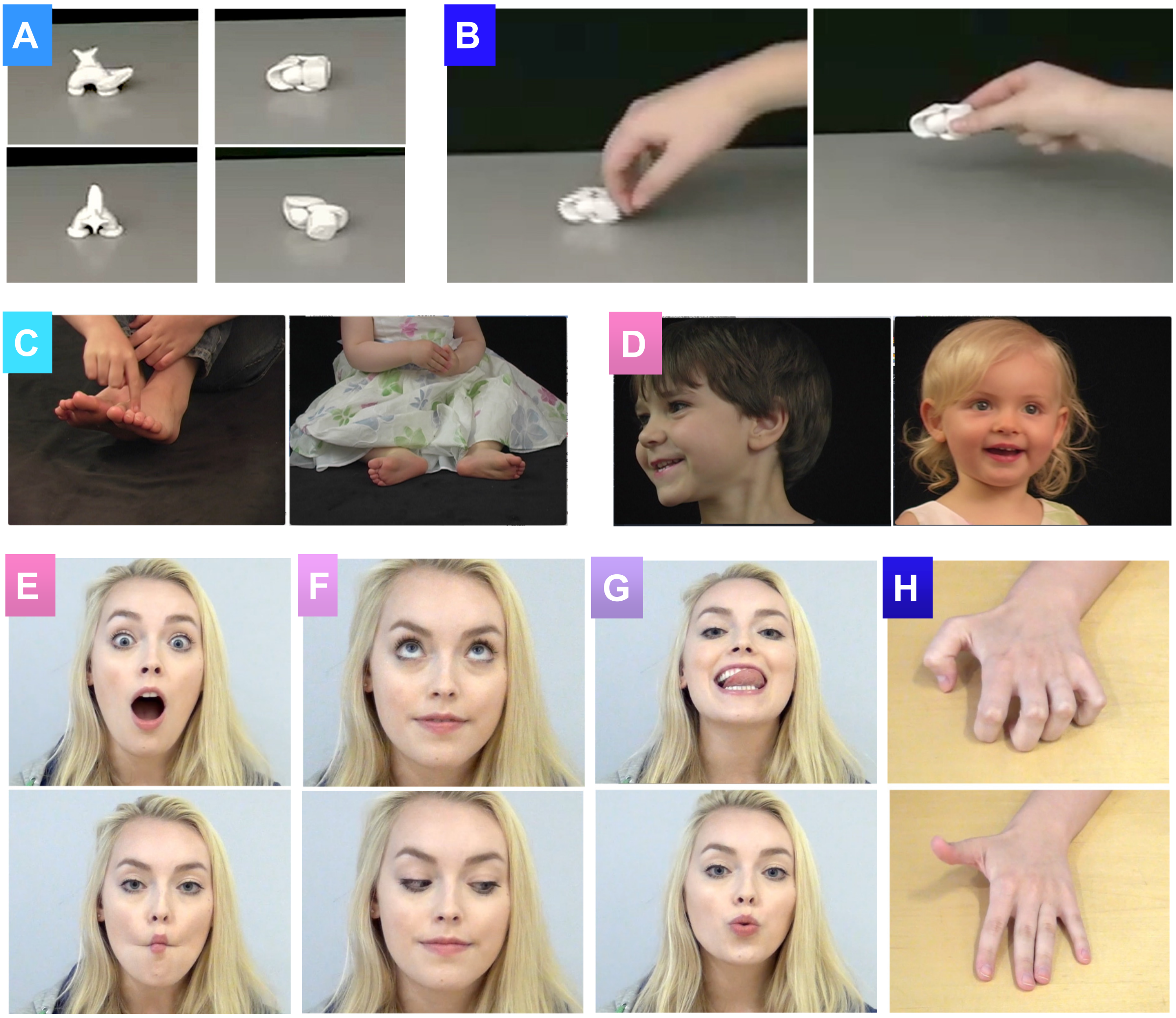

Fig. 1. Sample stimuli for each experiment and condition. *Experiment 1*: *A*: example objects, grouped vertically by family; *B*: example family of dissimilar actions. *Experiment 2*: *C*: example body action stimuli; *D*: example face action stimuli. *Experiment 3a/b*: *E*: example face actions; *F*: example eye actions; *G*: example mouth actions; *H*: example hand actions.

Experiment 3a/b: Face, Eye, Mouth, and Hand Action Observation/Imitation

Participants watched silent videos of an actress performing face actions, eye actions, mouth actions, and hand actions. Additionally, the experiment included videos where the actress pronounced consonant and vowel sounds, syllables, nonwords, and words in English and German, and sang or hummed nonwords, all in a blocked design. For the purposes of the current study, we examined the first four conditions: face actions, eye actions, mouth actions, and hand actions. In the observation version of the experiment (*experiment 3a*), participants were asked to just watch attentively, and in the imitation version (*experiment 3b*), a different set of participants were instructed to imitate each action while keeping their head as still as possible.

Materials. There were eight unique stimuli per condition. The videos depicted a female actress against a gray background. In the face, eye, and mouth action conditions, she was sitting facing the camera, with the frame going from just below her shoulders to just above the top of her head. Each video started and ended with the actress looking at the camera, with a neutral expression. The face condition included actions like looking surprised or making a "fish" face (see Fig. 1 for screenshots from sample stimuli); the eye condition included actions like moving the eyes up or to the lower left; and the mouth condition included actions like touching the upper teeth with the tongue or pursing the lips to blow air out. In the hand action condition, the hand rested on a wooden table, with the frame covering the hand and a portion of the forearm. Each video started and ended with the hand resting on the table. The hand condition included actions like pulling in the fingers or tapping a finger or multiple fingers on the table.

Procedure: experiment 3a (observation). Each video (trial) lasted 5 s, and trials were grouped into blocks of 3 trials each. Each run consisted of 26 experimental blocks (15 s each; 2 blocks for each of thirteen conditions) and 5 fixation blocks (14 s each), placed at the beginning and end of each run, as well as after the 7th, 13th, and 20th blocks. Each run thus lasted 460 s (7 min 30 s). Each participant saw between four and six runs.

Procedure: experiment 3b (imitation). The procedure was identical to that of *experiment 3a* except that each video (trial) lasted 8 s (5 s for the video and 3 s for the participant to imitate the action; note that although the videos lasted 5 s each, the actual action does not take longer than ~3 s). Each run thus lasted 694 s (11 min 34 s). Each participant saw between 3 and 8 runs.

fMRI Data Acquisition and Preprocessing

Structural and functional data were collected on the whole-body 3 Tesla Siemens Trio scanner with a 32-channel head coil at the

Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. T1-weighted structural images were collected in 128 axial slices with 1-mm isotropic voxels (TR = 2530 ms, TE = 3.48 ms). Functional, blood oxygenation level-dependent data were acquired using an EPI sequence (with a 90º degree flip angle and using GRAPPA with an acceleration factor of 2), with the following acquisition parameters: thirty-one 4-mm thick near-axial slices, acquired in an interleaved order with a 10% distance factor, 2.1 × 2.1 mm in-plane resolution; field of view of 200 ms in the phase encoding anterior to posterior (A > P) direction; matrix size of 96 × 96 mm; TR of 2,000 ms; and TE of 30 ms. Prospective acquisition correction (Thesen et al. 2000) was used to adjust the positions of the gradients based on the participant's motion from the previous TR. The first 10 s of each run (before the start of presentation of the stimuli) were excluded to allow for steady-state magnetization.

MRI data were analyzed using SPM5 and custom MATLAB and shell scripts. Each participant's data were motion corrected, normalized into a common brain space (Montreal Neurological Institute), and resampled into 2-mm isotropic voxels. The data were smoothed with a 4-mm Gaussian filter and high-pass filtered (at 200s). All task effects were estimated using a general linear model in which each experimental condition was modeled with a boxcar function convolved with the canonical hemodynamic response function.

Definition of Group-Constrained, Subject-Specific Functional Regions of Interest

The critical analyses were restricted to individually defined language functional regions of interest (fROIs). These fROIs were defined using the group-constrained subject-specific (GSS) approach (Fedorenko et al. 2010; Julian et al. 2012) where a set of spatial parcels (binary masks that correspond to locations where activation has been previously observed for the relevant localizer contrast) is combined with each individual subject's localizer activation map to constrain the definition of individual fROIs. The parcels are sufficiently large to encompass the extent of variability in the locations of individual activations. For the critical language fROIs, we used a set of six parcels derived from a group-level probabilistic activation overlap map for the sentences > nonwords contrast in 220 participants. These parcels (Fig. 2) included three regions in the left frontal cortex: two in the left inferior frontal gyrus (LIFG, LIFGorb), and one in the left middle frontal gyrus (LMFG), two in the left temporal lobe (LAntTemp and LPostTemp), and one extending into the angular gyrus (LAngG). These parcels are similar to the ones originally reported in Fedorenko et al. (2010) based on a probabilistic activation overlap map from 25 participants, except that the two anterior temporal parcels were grouped together (the original LAntTemp merged with LMidAntTemp), and the two posterior temporal parcels were grouped together (the original LMidPostTemp merged with LPostTemp). The parcels are available for download from https://evlab.mit.edu/funcloc.

Within each parcel, we selected the top 10% most responsive voxels, based on the *t* values for the sentences > nonwords contrast (see e.g., Fig. 1 in Blank et al. 2014; or Fig. 1 in Mahowald and Fedorenko 2016, for sample fROIs). Statistical tests were performed on these values.

In addition to the language fROIs, a set of control fROIs was defined in the participants in *experiments* 2, *3a*, and *3b*. In particular, we used 18 anatomical parcels across the two hemispheres (Tzourio-Mazoyer et al. 2002) covering frontal and parietal brain areas that belong to the so-called MD network (Duncan 2010, 2013). This network has been linked to executive demands across domains (e.g., Duncan and Owen 2000; Fedorenko et al. 2013; Hugdahl et al. 2015), but parts of this network have also been implicated in the processing of actions (e.g., Biagi et al. 2016; Caspers et al. 2010; Culham and Valyear 2006; Gallivan and Culham 2015). We thus expected some of these regions to respond to action observation and/or imitation. In particular, we focused on a subset of six parcels (although the results were corrected for the total number of regions that comprise this network, i.e., 18): the bilateral IFGop and PrecG fROIs in the frontal cortex because those lie in close proximity to the language fROIs, and the bilateral SupPar fROIs in the parietal cortex because these regions have been implicated in action observation in prior work (e.g., Johnson-Frey et al. 2005).

To define individual MD fROIs, we used a spatial working memory task where participants keep track of locations within a grid (this MD localizer task was not included in *experiment 1*; hence this analysis could not be performed for those participants). The task is described in detail in Fedorenko et al. (2013; see also Blank et al. 2014). Briefly, on each trial, participants saw a 3 × 4 grid and kept track of eight (hard version) or four (easy version) locations that were sequentially flashed two at a time or one at a time, respectively. Then, participants indicated their memory for these locations in a two-alternative, forced-choice paradigm via a button press. Feedback was provided after every trial. Hard and easy conditions were presented in a standard blocked design (4 trials in a 32-s block, 6 blocks per condition per run) with a counterbalanced order across runs. Each run included four blocks of fixation (16 s each) and lasted a total of 448 s. Within each anatomical parcel, we selected the top 10% most responsive voxels, based on the *t* values for the hard > easy spatial working memory contrast. Statistical tests were performed on these values.

Finally, for some additional analyses reported in the DISCUSSION, we examined *1*) brain regions in the auditory cortex that support speech perception, and *2*) brain regions in the premotor cortex that support speech articulation. For the former, we used the following anatomical parcels from the FSL atlas (Desikan et al. 2006): bilateral planum polare (PP), planum temporale (PT), anterior superior temporal gyrus (ASTG), and posterior superior temporal gyrus (PSTG). To define individual speech-responsive fROIs, these anatomical parcels were masked with activation maps for a contrast between listening to nonwords and observing hand actions (in *experiment 3a*). The responses were then extracted to nonwords and the four action observation conditions. To estimate the responses to the nonwords and hand action observation conditions, an across-runs cross-validation procedure was used so that the data to define the fROIs and estimate their responses were independent (e.g., Kriegeskorte 2011). In particular, all but one run were used to define the fROIs and the responses were estimated in the left-out run; this procedure was repeated leaving out each run in turn; the response estimates were then averaged across runs to derive a single estimate per condition per fROI. This procedure allows all of the data to be used while maintaining the independence between the data used to define the fROIs and the data used to examine their responses (e.g., Nieto-Castañón and Fedorenko 2012).

For the articulation regions, we used functional parcels derived from a group-level probabilistic activation overlap map for the contrast between the production of difficult-to-articulate nonwords and fixation in 20 participants, as reported in Basilakos et al. (2017). We focused on the regions in the premotor cortex bilaterally: a region in the left precentral gyrus, and two regions in the right precentral gyrus (see Fig. 3 in Basilakos et al. 2017). To define individual articulation-responsive fROIs, these parcels were masked with activation maps for a contrast between imitating nonwords (repeating the nonword produced by the actress) and imitating hand actions (in *experiment 3b*). The responses were then extracted to nonwords and the four action imitation conditions. As with the analyses of the speech-responsive regions, to estimate the responses to the nonwords and hand action imitation conditions, an across-runs cross-validation procedure was used so that the data to define the fROIs and estimate their responses were independent (e.g., Kriegeskorte 2011).

Analyses

In the critical analyses that examined the responses of the language fROIs to the different action observation/imitation conditions, we used two-tailed *t*-tests to compare the responses to each action condition against *1*) the low-level fixation baseline, *2*) nonword

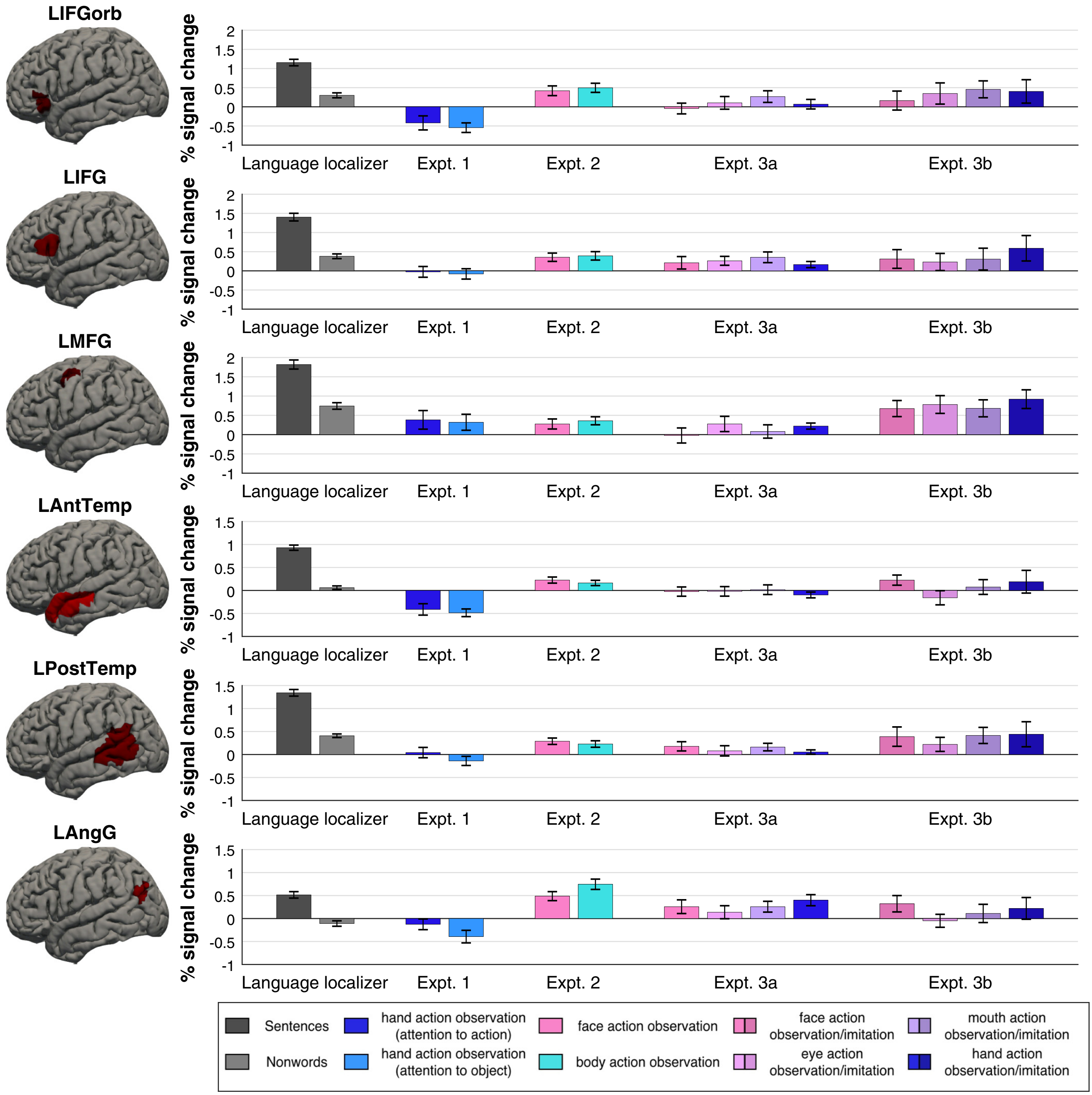

Fig. 2. Response to the language localizer conditions [estimated in data not used for functional region of interest (fROI) definition, as described in MATERIALS AND METHODS] and action conditions across experiments. Next to each bar graph, we show the language parcel used to constrain the selection of individual language fROIs; the individual fROIs constitute 10% of each parcel (see MATERIALS AND METHODS for details). Error bars indicate SE of the mean over participants. LIFG and LIFGorb, regions in left inferior frontal gyrus; LMFG, region in the left middle frontal gyrus; LAntTemp and LPostTemp, regions in the left temporal lobe; LAngG, region extending into the angular gyrus.

processing, which serves as the control condition in the language localizer, and *3*) sentence comprehension. The resulting *P* values were corrected for the number of language fROIs within each experiment (i.e., 6), using the false discovery rate correction (Benjamini and Yekutieli 2001). If language comprehension and action observation/imitation share computational demands, then the action conditions should elicit a response that is as strong as the sentence comprehension condition or, at least, reliably stronger than the nonword processing condition.

RESULTS

Behavioral Data

Overt behavioral responses were only collected in *experiment 1*, where participants watched videos and performed a one-back task on the action or the object in the video, as described in MATERIALS AND METHODS. Accuracies were high in both conditions but slightly and reliably higher for the actions condition than the objects condition [94.9 and 87.5%, respectively; two-tailed $t(9) = 3.18$, $P < 0.05$]. Furthermore, as expected (given that actions take time to unfold), participants were faster in the objects condition than the actions condition [1.37 vs. 1.71 s; two-tailed $t(9) = 6.05$, $P < = 0.0005$].

Validation of the Language fROIs

Replicating previous work (Fedorenko et al. 2010, 2011), the sentences > nonwords effect was highly reliable in each of six fROIs both *1*) across the entire set of participants [$t(76) > 10$, $P < 0.0001$], and *2*) in each experiment individually [*experiment 1*: $t(9) > 4.43$, $P < 0.001$, *experiment 2*: $t(53) > 8.39$, $P < 0.0001$, *experiment 3a*: $t(12) > 3.68$, $P < 0.005$, and *experiment 3b*: $t(12) > 4.01$, $P < 0.001$]. Here, and in validating the MD fROIs, an across-runs cross-validation procedure, described above, was used so that data used to define the fROIs were independent of the data used to estimate the responses.

Reponses of the Language fROIs to the Action Conditions

The results are reported in Table 2 and Fig. 2. Across experiments, none of the language regions responded strongly and consistently to action observation or imitation. In most fROIs, the action conditions failed to elicit a response above the fixation baseline (except for *experiment 2*, where both conditions elicited small but reliable above-baseline responses in all language fROIs). Furthermore, the response to the action observation/imitation condition did not significantly differ from the nonword condition, with the exception of the AngG fROI, which responded more strongly to some action observation conditions than the nonword condition. Finally, again with the exception of the AngG fROI, the response to the action observation/imitation condition was reliably (in almost all cases, and always numerically) below that elicited by sentence comprehension.

Experiment 1. When participants watched videos of a hand performing simple manipulations of an object, there was no above-baseline response in any of the language fROIs, regardless of whether participants were asked to focus on the objects [$t(9) < 1.5$, NS] or actions [$t(9) < 1.6$, NS]. Furthermore, neither of the action conditions elicited a response that was reliably greater than the nonword condition, whereas the sentence condition elicited a reliably greater response than either of the two action conditions [$t(9) > 2.75$, $P < 0.05$.

Experiment 2. In this experiment, every language fROI showed a reliably above-baseline response to both the face action observation condition [$t(53) > 2.11$, $P < 0.05$] and the body action observation condition [$t(53) > 2.86$, $P < 0.01$]. However, in all fROIs except for the AngG fROI, this response was *1*) not reliably higher than that elicited by the nonword condition [$t(53) < 1.67$, $P > 0.16$], and *2*) reliably lower than that elicited by the sentence condition [$t(53) > 5.72$, $P < 0.0001$]. In the AngG language fROI, both action observation conditions elicited a response that was reliably stronger than that elicited by the nonword condition and that did not differ from that elicited by the sentence condition. We come back to the AngG fROI in the DISCUSSION.

Experiment 3a. Similar to *experiment 1*, there was no above-baseline response in the language fROIs to any of the four conditions, with the exception of the AngG fROI and the MFG fROI, which showed reliably above-baseline responses to hand action observation [$t(12) > 2.82$, $P < 0.05$], but only the AngG fROI responded reliably more strongly to hand action observation (and mouth action observation) than to nonwords [$t(12) > 3.67$, $P < 0.05$]; in all other fROIs, none of the action observation conditions produced a stronger response than nonwords. Finally, in all language fROIs, except for the AngG fROI, the sentence condition elicited a reliably greater response than each of the four action observation conditions [$t(12) > 3.30$, $P < 0.01$]. In the AngG fROI, the response to the action observation conditions did not reliably differ in magnitude from the sentence condition.

Experiment 3b. In this experiment, where participants observed and imitated different kinds of actions, there was no above-baseline responses except for the MFG fROI, which responded reliably above baseline to the eye, mouth, and hand action conditions [$t(12) > 2.23$, $P < 0.05$] and marginally to the face action condition [$t(12) = 3.09$, $P = 0.056$]. However, these responses did not significantly differ from the response elicited by the nonword condition (see Fedorenko et al. 2011, for a similar pattern of results with other nonlinguistic tasks). Furthermore, the sentence condition elicited a reliably or marginally greater response than each of the four action conditions in all language fROIs, except for the AngG fROI and some frontal fROIs for some of the conditions (see Table 2 for details).

Validation of the Control (MD) fROIs

Replicating previous work (Blank et al. 2014; Fedorenko et al. 2013), the hard > easy spatial working memory effect was highly reliable in each of six fROIs across participants with 2 runs [$t(47) > 7.8$, $P < 0.0001$]. Participants with one run only ($n = 18$) could not be included in this validation analysis because across-runs cross-validation could not be performed; for those participants, we ensured that MD activations looked as expected based on visual examination of whole brain activation maps.

Responses of the Control (MD) fROIs to the Action Conditions

Unlike in the language fROIs, all action imitation conditions elicited reliably above-baseline responses in almost all MD fROIs. Similarly, at least some action observation conditions elicited reliable responses. The body action observation condition from *experiment 2* and the eye and hand action observation conditions from *experiment 3a* elicited the strongest responses. Strong responses to eye movement observation and imitation could be related to prior claims about the role of this fronto-parietal system in saccades (e.g., Pierrot-Deseilligny et al. 2004). Results are reported in Table 3 and Fig. 3.

Responses of Speech Perception and Articulation Regions to the Action Conditions

As discussed at the beginning of MATERIALS AND METHODS, we have here focused on high-level language processing regions (e.g., Fedorenko et al. 2010), which plausibly store our linguistic knowledge that we use to both interpret and generate meaningful utterances (e.g., Menenti et al. 2011). These regions are distinct from lower level speech perception regions (e.g., Norman-Haignere et al. 2015; Overath et al. 2015) and

Table 2. *Results for each experiment and condition for the six language fROIs*

Experiment/Condition/ROI	Action Condition vs. Fixation	Action Condition vs. Nonwords	Action Condition vs. Sentences
Experiment 1			
Hand action observation (attention to action)			
LIFGorb	$t(9) = -2.29, P = 0.143$	$t(9) = -1.33, P = 0.357$	$t(9) = 4.18, P < 0.005$
LIFG	$t(9) < \|1\|$, n.s.	$t(9) < \|1\|$, n.s.	$t(9) = 3.31, P < 0.05$
LMFG	$t(9) = 1.59, P = 0.294$	$t(9) = -1.66, P = 0.357$	$t(9) = 5.67, P < 0.005$
LAntTemp	$t(9) = -3.32, P = 0.053$	$t(9) < \|1\|$, n.s.	$t(9) = 3.88, P < 0.01$
LPostTemp	$t(9) < \|1\|$, n.s.	$t(9) = -1.26, P = 0.357$	$t(9) = 4.65, P < 0.005$
LAngG	$t(9) = -1.08, P = 0.463$	$t(9) = 2.33, P = 0.271$	$t(9) = 2.76, P < 0.05$
Hand action observation (attention to object)			
LIFGorb	$t(9) = -4.35, P < 0.01$	$t(9) = -2.10, P = 0.130$	$t(9) = 5.42, P < 0.001$
LIFG	$t(9) < \|1\|$, n.s.	$t(9) = -1.41, P = 0.288$	$t(9) = 4.24, P < 0.005$
LMFG	$t(9) = 1.54, P = 0.234$	$t(9) = -2.45, P = 0.110$	$t(9) = 5.82, P < 0.001$
LAn*t*Temp	$t(9) = -5.82, P < 0.005$	$t(9) = -1.16, P = 0.291$	$t(9) = 4.72, P < 0.005$
LPostTemp	$t(9) = -1.40, P = 0.234$	$t(9) = -2.52, P = 0.110$	$t(9) = 6.10, P < 0.001$
LAngG	$t(9) = -2.87, P < 0.05$	$t(9) = 1.12, P = 0.291$	$t(9) = 4.75, P < 0.005$
Experiment 2			
Face action observation			
LIFGorb	$t(53) = 3.30, P < 0.005$	$t(53) < \|1\|$, n.s.	$t(53) = 6.19, P < 0.0001$
LIFG	$t(53) = 3.28, P < 0.005$	$t(53) < \|1\|$, n.s.	$t(53) = 7.98, P < 0.0001$
LMFG	$t(53) = 2.12, P < 0.05$	$t(53) = -3.48, P < 0.005$	$t(53) = 9.36, P < 0.0001$
LantTemp	$t(53) = 3.41, P < 0.005$	$t(53) = 1.66, P = 0.157$	$t(53) = 8.89, P < 0.0001$
LpostTemp	$t(53) = 4.14, P < 0.0005$	$t(53) = -1.65, P = 0.157$	$t(53) = 9.69, P < 0.0001$
LangG	$t(53) = 4.98, P < 0.0001$	$t(53) = 4.52, P < 0.0005$	$t(53) < \|1\|$, n.s.
Body action observation			
LIFGorb	$t(53) = 4.16, P < 0.0005$	$t(53) < \|1\|$, n.s.	$t(53) = 5.73, P < 0.0001$
LIFG	$t(53) = 3.56, P < 0.005$	$t(53) < \|1\|$, n.s.	$t(53) = 7.76, P < 0.0001$
LMFG	$t(53) = 3.46, P < 0.005$	$t(53) = -3.57, P < 0.005$	$t(53) = 10.50, P < 0.0001$
LantTemp	$t(53) = 2.87, P < 0.01$	$t(53) < \|1\|$, n.s.	$t(53) = 9.64, P < 0.0001$
LpostTemp	$t(53) = 3.23, P < 0.005$	$t(53) = -2.22, P = 0.061$	$t(53) = 10.31, P < 0.0001$
LangG	$t(53) = 6.66, P < 0.0001$	$t(53) = 6.00, P < 0.0001$	$t(53) = -1.83, P = 0.073$
Experiment 3a			
Face action observation			
LIFGorb	$t(12) = 1.76, P = 0.156$	$t(12) < \|1\|$, n.s.	$t(12) = 4.74, P < 0.001$
LIFG	$t(12) = 2.53, P = 0.146$	$t(12) < \|1\|$, n.s.	$t(12) = 3.31, P < 0.01$
LMFG	$t(12) < \|1\|$, n.s.	$t(12) = -2.11, P = 0.169$	$t(12) = 5.17, P < 0.001$
LAntTemp	$t(12) < \|1\|$, n.s.	$t(12) < \|1\|$, n.s.	$t(12) = 5.24, P < 0.001$
LPostTemp	$t(12) = 1.97, P = 0.146$	$t(12) = -1.10, P = 0.582$	$t(12) = 4.78, P < 0.001$
LAngG	$t(12) = 2.19, P = 0.146$	$t(12) = 2.65, P = 0.126$	$t(12) = 1.13, P = 0.282$
Eye action observation			
LIFGorb	$t(12) < \|1\|$, n.s.	$t(12) < \|1\|$, n.s.	$t(12) = 4.03, P < 0.005$
LIFG	$t(12) = 2.25, P = 0.263$	$t(12) < \|1\|$, n.s.	$t(12) = 3.64, P < 0.005$
LMFG	$t(12) = 1.40, P = 0.562$	$t(12) = -1.20, P = 0.509$	$t(12) = 4.33, P < 0.005$
LAntTemp	$t(12) < \|1\|$, n.s.	$t(12) < \|1\|$, n.s.	$t(12) = 6.95, P < 0.0001$
LPostTemp	$t(12) < \|1\|$, n.s.	$t(12) = -1.37, P = 0.509$	$t(12) = 5.20, P < 0.001$
LAngG	$t(12) < \|1\|$, n.s.	$t(12) = 1.66, P = 0.509$	$t(12) = 1.62, P = 0.131$
Mouth action observation			
LIFGorb	$t(12) < \|1\|$, n.s.	$t(12) = -2.11, P = 0.113$	$t(12) = 5.72, P < 0.0005$
LIFG	$t(12) = 1.29, P = 0.440$	$t(12) < \|1\|$, n.s.	$t(12) = 4.10, P < 0.005$
LMFG	$t(12) < \|1\|$, n.s.	$t(12) = -2.54, P = 0.078$	$t(12) = 5.11, P < 0.001$
LAntTemp	$t(12) < \|1\|$, n.s.	$t(12) < \|1\|$, n.s.	$t(12) = 6.93, P < 0.0001$
LPostTemp	$t(12) = 1.77, P = 0.328$	$t(12) = -1.20, P = 0.383$	$t(12) = 4.69, P < 0.001$
LAngG	$t(12) = 1.73, P = 0.328$	$t(12) = 3.68, P < 0.05$	$t(12) = 1.04, P = 0.318$
Hand action observation			
LIFGorb	$t(12) < \|1\|$, n.s.	$t(12) = -1.14, P = 0.353$	$t(12) = 4.80, P < 0.001$
LIFG	$t(12) = 2.04, P = 0.127$	$t(12) < \|1\|$, n.s.	$t(12) = 4.24, P < 0.005$
LMFG	$t(12) = 2.83, P < 0.05$	$t(12) = -1.81, P = 0.190$	$t(12) = 5.13, P < 0.0005$
LAntTemp	$t(12) = -1.55, P = 0.220$	$t(12) = -1.10, P = 0.353$	$t(12) = 5.84, P < 0.0005$
LPostTemp	$t(12) = 1.15, P = 0.326$	$t(12) = -2.92, P < 0.05$	$t(12) = 6.48, P < 0.0005$
LAngG	$t(12) = 3.31, P < 0.05$	$t(12) = 4.38, P < 0.01$	$t(12) < \|1\|$, n.s.
Experiment 3b			
Face action imitation			
LIFGorb	$t(12) = 2.06, P = 0.123$	$t(12) < \|1\|$, n.s.	$t(12) = 2.06, P = 0.075$
LIFG	$t(12) = 1.08, P = 0.455$	$t(12) < \|1\|$, n.s.	$t(12) = 3.27, P < 0.05$
LMFG	$t(12) = 3.09, P = 0.056$	$t(12) = -2.11, P = 0.169$	$t(12) = 2.52, P < 0.05$
LAntTemp	$t(12) < \|1\|$, n.s.	$t(12) < \|1\|$, n.s.	$t(12) = 4.62, P < 0.005$
LPostTemp	$t(12) = 2.37, P = 0.107$	$t(12) = -1.10, P = 0.582$	$t(12) = 4.10, P < 0.005$
LAngG	$t(12) < \|1\|$, n.s.	$t(12) = 2.65, P = 0.126$	$t(12) = 1.77, P = 0.101$

Continued

Table 2.—*Continued*

Experiment/Condition/ROI	Action Condition vs. Fixation	Action Condition vs. Nonwords	Action Condition vs. Sentences
Eye action imitation			
LIFGorb	$t(12) = 1.26, P = 0.386$	$t(12) < \|1\|$, n.s.	$t(12) = 1.78, P = 0.101$
LIFG	$t(12) = 1.03, P = 0.386$	$t(12) < \|1\|$, n.s.	$t(12) = 3.17, P < 0.05$
LMFG	$t(12) = 3.39, P < 0.05$	$t(12) = -1.47, P = 0.335$	$t(12) = 2.25, P = 0.053$
LAntTemp	$t(12) = -1.04, P = 0.386$	$t(12) = -2.17, P = 0.304$	$t(12) = 6.06, P < 0.0005$
LPostTemp	$t(12) = 1.44, P = 0.386$	$t(12) = -1.7, P = 0.335$	$t(12) = 5.16, P < 0.001$
LAngG	$t(12) < \|1\|$, n.s.	$t(12) < \|1\|$, n.s.	$t(12) = 3.23, P < 0.05$
Mouth action imitation			
LIFGorb	$t(12) < \|1\|$, n.s.	$t(12) < \|1\|$, n.s.	$t(12) = 2.84, P < 0.05$
LIFG	$t(12) = 1.27, P = 0.274$	$t(12) < \|1\|$, n.s.	$t(12) = 3.01, P < 0.05$
LMFG	$t(12) = 3.24, P < 0.05$	$t(12) < \|1\|$, n.s.	$t(12) = 2.52, P < 0.05$
LAntTemp	$t(12) = 2.04, P = 0.144$	$t(12) < \|1\|$, n.s.	$t(12) = 7.24, P < 0.0001$
LPostTemp	$t(12) = 1.85, P = 0.144$	$t(12) < \|1\|$, n.s.	$t(12) = 5.10, P < 0.001$
LAngG	$t(12) = 1.81, P = 0.144$	$t(12) = 1.89, P = 0.495$	$t(12) = 1.22, P = 0.247$
Hand action imitation			
LIFGorb	$t(12) = 1.32, P = 0.319$	$t(12) < \|1\|$, n.s.	$t(12) = 1.52, P = 0.163$
LIFG	$t(12) = 1.78, P = 0.263$	$t(12) = 1.15, P = 0.549$	$t(12) = 1.77, P = 0.163$
LMFG	$t(12) = 3.80, P < 0.05$	$t(12) = 1.34, P = 0.549$	$t(12) = 1.49, P = 0.163$
LAntTemp	$t(12) < \|1\|$, n.s.	$t(12) < \|1\|$, n.s.	$t(12) = 4.76, P < 0.005$
LPostTemp	$t(12) = 1.62, P = 0.263$	$t(12) < \|1\|$, n.s.	$t(12) = 4.16, P < 0.005$
LAngG	$t(12) < \|1\|$, n.s.	$t(12) = 1.23, P = 0.549$	$t(12) = 1.49, P = 0.163$

Note that although in Fig. 2, we plot the language localizer responses across the entire set of participants in the current study, all the comparisons between the action conditions and the conditions of the language localizer experiment were performed within each experiment separately using two-tailed paired-samples *t*-tests. In *columns 2* and *3*, significance values indicate whether the action observation/imitation condition elicited a response reliably above the baseline (*column 2*) or reliably above the nonword condition (*column 3*). In *column 4*, significance values indicate whether the action observation/imitation condition elicited a response reliably below the sentence condition. LIFG and LIFGorb, regions in left inferior frontal gyrus; LMFG, region in the left middle frontal gyrus; LAntTemp and LpostTemp, regions in the left temporal lobe; LangG, region extending into the angular gyrus; ROI, region of interest.

from speech articulation regions (e.g., Basilakos et al. 2017; Bohland and Guenther 2006; Flinker et al. 2015). Might some of this perceptual or motor speech machinery overlap with action observation or imitation? Based on the available evidence, a tentative answer appears to be "no." In particular, the superior temporal regions that respond robustly to speech show some response during speech articulation (e.g., Basilakos et al. 2017; Hickok et al. 2009) but respond very little when participants produce even actions that involve speech articulators, i.e., nonspeech oral-motor movements (Basilakos et al. 2017). To shed further light on this question, we performed an additional analysis on data from *experiment 3a*. We used a contrast between listening to nonwords and hand action observation to define speech-responsive regions within the superior temporal cortex and then examined the responses of those regions to nonwords and hand action observation (in data not used for fROI definition), as well as to face, eye, and mouth action observation conditions. As Fig. 4*A* clearly shows, the four action observation conditions fail to elicit above-baseline responses, suggesting that these regions do not support action observation.

What about regions that support speech articulation? Basilakos et al. (2017) report quite robust responses to the production of nonspeech oral-motor movements in premotor articulation regions. We performed an additional analysis on data from *experiment 3b* to examine the responses of those articulation regions to action imitation more broadly. We used a contrast between imitating nonwords (repeating the nonword produced by the actress) and hand actions to define articulation-responsive regions within ventral premotor cortex, and then examined the responses of those regions to nonwords and hand action imitation (in data not used for fROI definition), as well as to face, eye, and mouth action imitation. As Fig. 4*B* shows, the mouth action imitation condition elicits as strong a response as, or a stronger response than, articulation, replicating Basilakos et al. (2017). The face condition (which also includes some mouth movements) also elicits a strong response. However, the hand and eye action imitation conditions elicit much lower responses. This relative selectivity for speech and oral-motor/face actions is in line with the idea that these regions contain a map of our articulatory apparatus (e.g., Bouchard et al. 2013; Guenther 2016), arguing against broad engagement in action imitation, as well as with prior findings of somatotopic organization in the motor areas (e.g., D'Ausilio et al. 2009; Murakami et al. 2011; Pulvermüller et al. 2006; Watkins et al. 2003).

Thus, similar to high-level language processing regions, speech perception regions do not appear to support action observation, and speech articulation regions do not appear to support action imitation.

DISCUSSION

We asked whether any part(s) of the language network, a set of brain regions that support high-level language processing (e.g., Fedorenko et al. 2010; Fedorenko and Thompson-Schill 2014), respond to action observation and/or imitation. Neural machinery that supports both language processing and some aspects of action observation/imitation has been postulated based on two distinct ideas. First, inspired by the discovery of mirror neurons in macaques (Rizolatti and Arbib 1988), some have argued that manual actions served as a fundamental precursor to linguistic communication in the evolution of our species (e.g., Arbib 2005; but see, e.g., Corballis 2003; Tomasello 2008, for arguments for gesture-based origins of language that do not hinge on the mirror-neuron-based theorizing). Second, some have postulated an amodal hierarchical proces-

Table 3. *Results for each experiment (for experiments 2 and 3a/b; no MD localizer was included in experiment 1) and condition for the six MD fROIs*

Experiment/Condition/ROI	Action Condition vs. Fixation
Experiment 2	
Face action observation	
LIFGop	$t(54) < \|1\|$, n.s.
RIFGop	$t(54) = 1.15, P = 0.521$
LPrecG	$t(54) < \|1\|$, n.s.
RPrecG	$t(54) = 2.25, P = 0.169$
LParSup	$t(54) = -1.66, P = 0.449$
RParSup	$t(54) = -1.28, P = 0.521$
Body action observation	
LIFGop	$t(54) = 1.58, P = 0.197$
RIFGop	$t(54) = 1.89, P = 0.521$
LPrecG	$t(54) = 3.06, P < 0.05$
RPrecG	$t(54) = 4.51, P < 0.0005$
LParSup	$t(54) = 4.82, P < 0.0005$
RParSup	$t(54) = 4.22, P < 0.001$
Experiment 3a	
Face action observation	
LIFGop	$t(12) = 1.27, P = 0.516$
RIFGop	$t(12) = 1.46, P = 0.514$
LPrecG	$t(12) = 1.40, P = 0.514$
RPrecG	$t(12) = 1.97, P = 0.514$
LParSup	$t(12) = 1.36, P = 0.514$
RParSup	$t(12) < \|1\|$, n.s.
Eye action observation	
LIFGop	$t(12) = 2.52, P = 0.081$
RIFGop	$t(12) = 3.68, P < 0.05$
LPrecG	$t(12) = 3.40, P < 0.05$
RPrecG	$t(12) = 4.66, P < 0.01$
LParSup	$t(12) = 3.15, P < 0.05$
RParSup	$t(12) = 2.03, P = 0.141$
Mouth action observation	
LIFGop	$t(12) = 2.32, P = 0.349$
RIFGop	$t(12) = 1.40, P = 0.507$
LPrecG	$t(12) = 1.89, P = 0.499$
RPrecG	$t(12) = 2.47, P = 0.349$
LParSup	$t(12) = 1.37, P = 0.507$
RParSup	$t(12) < \|1\|$, n.s.
Hand action observation	
LIFGop	$t(12) = 2.00, P = 0.207$
RIFGop	$t(12) = 2.38, P = 0.126$
LPrecG	$t(12) = 2.67, P = 0.091$
RPrecG	$t(12) = 3.91, P < 0.05$
LParSup	$t(12) = 2.80, P = 0.091$
RParSup	$t(12) = 1.83, P = 0.236$
Experiment 3b	
Face action observation	
LIFGop	$t(12) = 2.87, P < 0.05$
RIFGop	$t(12) = 3.23, P < 0.05$
LPrecG	$t(12) = 4.50, P < 0.005$
RPrecG	$t(12) = 7.56, P < 0.0005$
LParSup	$t(12) = 5.23, P < 0.001$
RParSup	$t(12) = 3.29, P < 0.05$
Eye action observation	
LIFGop	$t(12) = 2.09, P = 0.117$
RIFGop	$t(12) = 2.71, P < 0.05$
LPrecG	$t(12) = 2.78, P < 0.05$
RPrecG	$t(12) = 3.64, P < 0.05$
LParSup	$t(12) = 3.05, P < 0.05$
RParSup	$t(12) = 1.80, P = 0.164$
Mouth action observation	
LIFGop	$t(12) = 3.97, P < 0.01$
RIFGop	$t(12) = 3.26, P < 0.05$
LPrecG	$t(12) = 4.43, P < 0.005$
RPrecG	$t(12) = 4.69, P < 0.005$
LParSup	$t(12) = 4.17, P < 0.005$
RParSup	$t(12) = 2.05, P = 0.088$
Hand action observation	
LIFGop	$t(12) = 3.38, P < 0.01$
RIFGop	$t(12) = 4.44, P < 0.005$
LPrecG	$t(12) = 4.50, P < 0.005$
RPrecG	$t(12) = 4.83, P < 0.005$
LParSup	$t(12) = 4.24, P < 0.005$
RParSup	$t(12) = 4.50, P < 0.005$

Significance values indicate whether the action observation/imitation condition elicited a response reliably above the baseline. LIFGop, RIFGop, LPrecG, and RprecG, regions in the frontal cortex; LSupPar and RSupPar, regions of the parietal cortex; n.s., not significant.

sor in the left frontal cortex (in or near "Broca's area") that is hypothesized to support both language processing and action perception/planning (e.g., Fiebach and Schubotz 2006; Koechlin and Jubault 2006; Tettamanti and Weniger 2006).

Across three experiments (77 participants, 90 scanning sessions), we examined neural responses of functionally defined language regions to a broad range of *action observation* conditions, including hand actions with (*experiment 1*) or without (*experiment 3a*) a manipulable object, but also actions that involve the face or face parts (*experiments 2* and *3a*) and body parts other than the hands (*experiment 2*). In the fourth experiment (13 participants), we further examined responses of language regions to *action imitation*, again involving different face and body parts.

The key result is that, with a single exception discussed below, none of the language regions responded strongly and consistently to action observation or imitation. In most language regions, the action conditions did not elicit a response above the fixation baseline, which suggests that the language regions are as active during action observation/imitation as they are when we are looking at a blank screen. Furthermore, in most language regions, the response to the action observation/imitation conditions *1*) did not significantly differ from the response elicited by the nonword condition (the control condition in the language localizer task), and *2*) was reliably lower than the response elicited by the sentence condition. These results suggest that language regions are selective for language processing, in line with earlier work that established selectivity for language relative to arithmetic, executive processing, music perception, and social cognition (e.g., Fedorenko and Varley 2016). This conclusion is also consistent with lesion studies that have reported dissociations between linguistic deficits and deficits in action observation/production (e.g., Sirigu et al. 1998) and with a recent fMRI study that showed that the degree of lateralization for language appears to be unrelated to the degree of lateralization for action observation (Häberling et al. 2016).

The only exception was the language fROI in the angular gyrus. This region responded more strongly to some action observation conditions than to nonwords, and, in some cases, the response to action observation was not significantly lower than the response to sentences. Evidence is accumulating that this region differs functionally from the rest of the language network. In particular, it shows relatively low functional correlations with other language regions during naturalistic cognition (e.g., Blank et al. 2014), including when using dynamic network modeling (Chai et al. 2016), and it shows lower

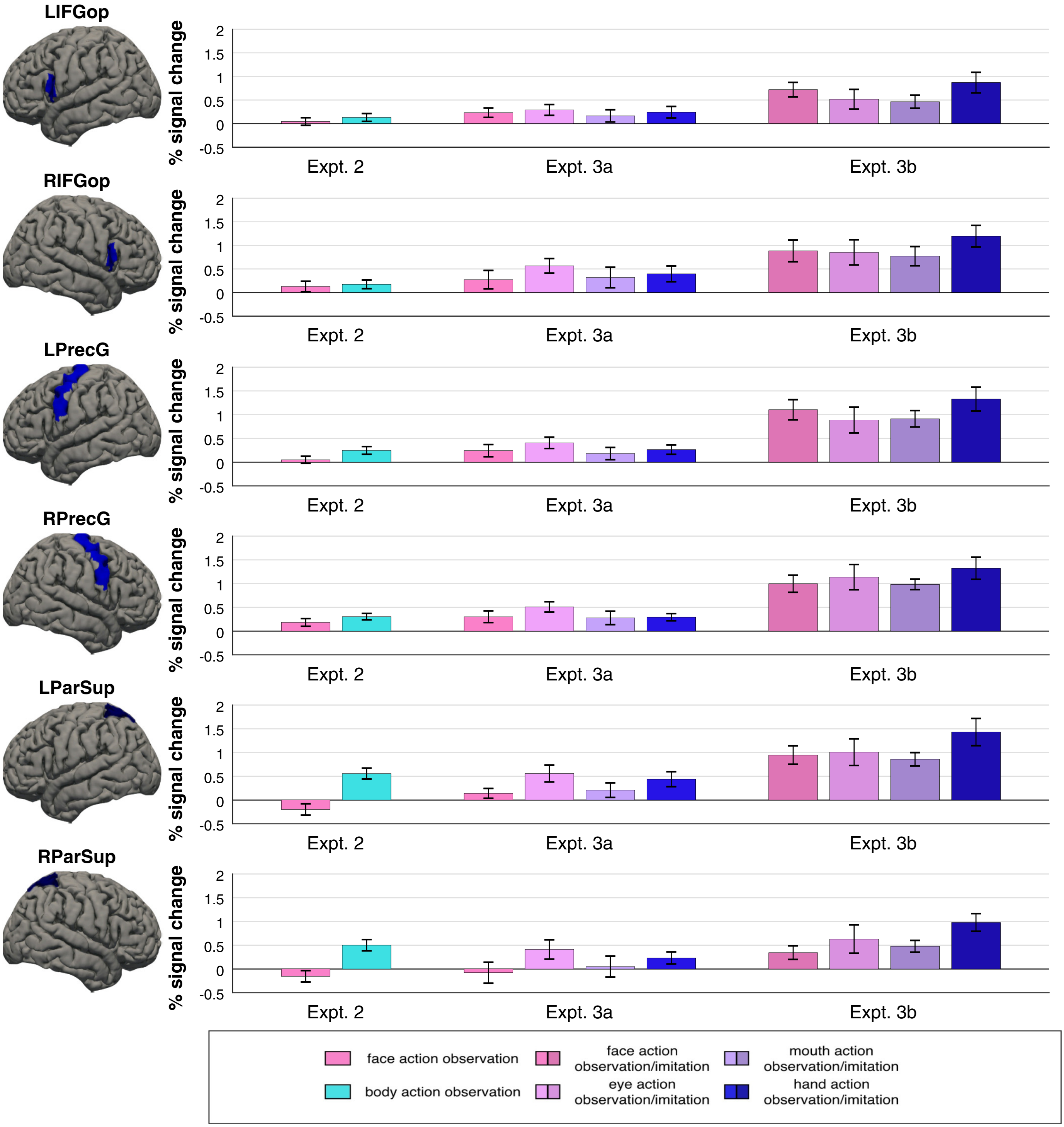

Fig. 3. Responses in multiple-demand (MD) regions to the action conditions in *experiments 2* and *3a/b*. Next to each bar graph, we show the MD parcels used to constrain the selection of individual MD functional regions of interest (fROIs); the individual fROIs constitute 10% of each parcel (see MATERIALS AND METHODS for details). Error bars indicate SE of the mean over participants. LIFGop, RIFGop, LPrecG, and RprecG, regions in the left and right inferior frontal gyrus and precentral gyrus; LSupPar and RSupPar, regions in the superior parietal cortex.

correlations in effect sizes and lateralization (e.g., Mahowald and Fedorenko 2016). It also differs from the other language regions in sensitivity to linguistic and nonlinguistic manipulations. For example, the AngG language fROI was the only region that did not show sensitivity to syntactic complexity (Blank et al. 2016), and it was the only region that did not respond more strongly to sentences than photographs matched for semantic content (Amit et al. 2017). The latter result suggests that the AngG language fROI may respond to visual stimuli in general, as opposed to action observation specifically. However, the precise role of this region in human cognition remains to be discovered. One current hypothesis (formulated not specifically about the language-responsive portion of the angular gyrus but about the broad anatomical area) is that it is “involved in all

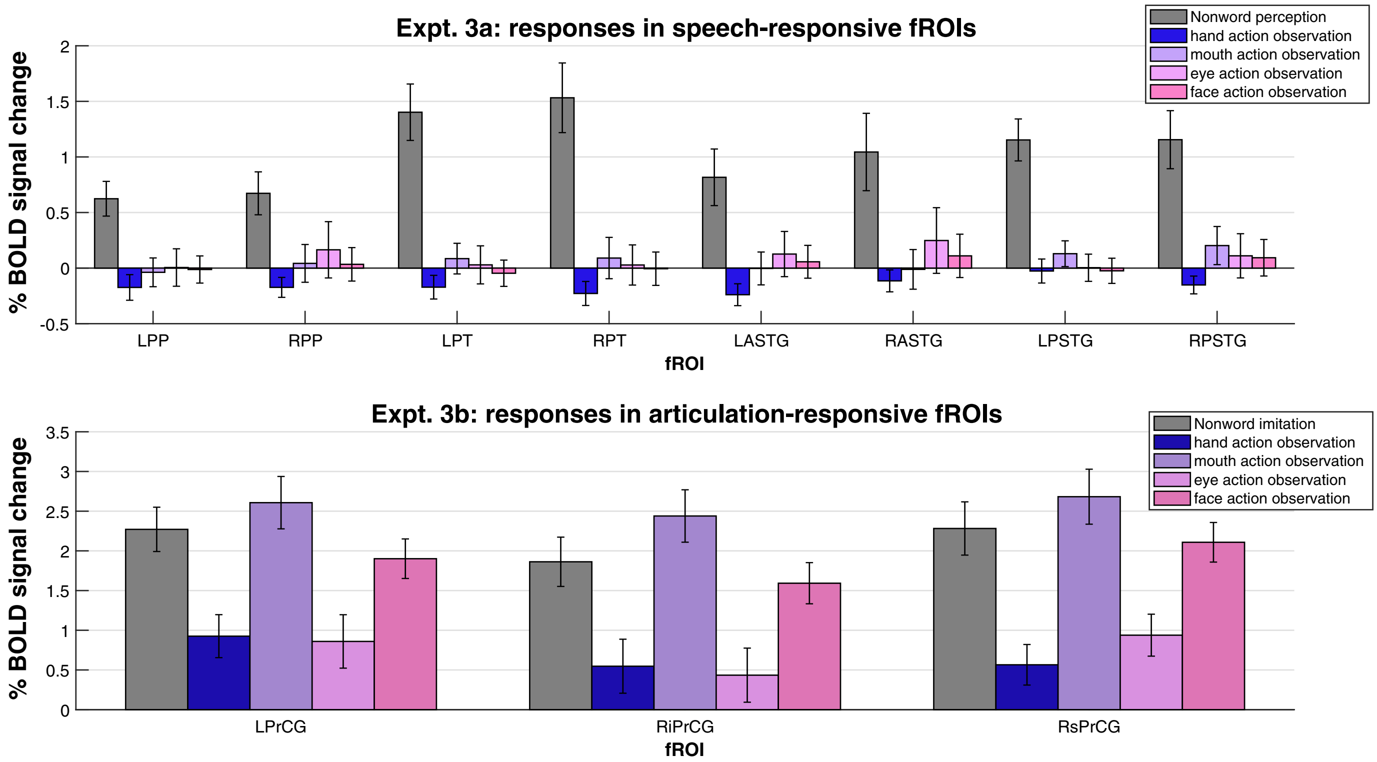

Fig. 4. *A*: responses in speech-responsive functional regions of interest (fROIs) in the auditory cortex (defined by nonword perception > hand action observation contrast in *experiment 3a*, see MATERIALS AND METHODS for details) to the nonword condition and the four action observation conditions in *experiment 3a*. Responses are estimated using data not used for fROI definition (see MATERIALS AND METHODS). Error bars indicate SE of the mean over participants. L/R PP, left/right planum polare; L/R PT, left/right planum temporale; L/R ASTG, left/right anterior superior temporal gyrus; L/R PSTG, left/right posterior superior temporal gyrus. *B*: responses in articulation-responsive fROIs in the premotor cortex (defined by the nonword imitation > hand action imitation contrast in *experiment 3b*, see MATERIALS AND METHODS for details) to the nonword imitation condition and the four action imitation conditions in *experiment 3b*. Responses are estimated using data not used for fROI definition (see MATERIALS AND METHODS). Error bars indicate SE of the mean over participants. LPrCG, left precentral gyrus; RiPrCG, right inferior precentral gyrus; RsPrCG, right superior precentral gyrus.

aspects of semantic processing" and contributes to "behaviors requiring fluent conceptual combination" (e.g., Binder et al. 2009; cf. Lambon-Ralph et al. 2017).

We now touch on four theoretical issues that the current results bear on.

Gestural Origins of Language

Just because in the modern human brain language processing and action observation appear to recruit nonoverlapping machinery does not imply that our linguistic communication system could not have arisen from the manual modality. In fact, this possibility is still perhaps the most plausible (e.g., Corballis 2003; Tomasello 2008; cf. Shepherd and Freiwald 2018; Slocombe 2015). However, once humans began to develop an extensive set of vocal communication signals, they plausibly had to allocate some portions of the association cortexes, massively expanded in the human brain (e.g., Buckner and Krienen 2013), to store these form-meaning mappings (see also Häberling et al. 2016). Given the differences between linguistic and nonlinguistic communication signals (including both discrete, categorical speech-accompanying gestures and continuous, mimetic facial expressions and body language), it is perhaps to be expected that these different forms of communication would recruit distinct cognitive (e.g., Goldin-Meadow and Brentari 2017; McNeill 1992) and neural (e.g., Häberling et al. 2016) machinery given the distinct computational demands they place on the mind and brain. It is worth noting that a few prior neuroimaging studies have argued that gesture processing does recruit the same brain regions as language comprehension (e.g., Andric et al. 2013; Enrici et al. 2011; Redcay et al. 2016; Villarreal et al. 2008; Xu et al. 2009; see Marstaller and Burianová 2014; Willems and Hagoort 2007; Yang et al. 2015, for reviews). However, those studies typically used symbolic gestures, pantomime, or "emblems" (e.g., wave, hold out hand for a shake, etc.). Given that such gestures are clearly associated with particular meanings, their processing may lead to the activation of the corresponding linguistic representations. Thus the overlap may be explained by the engagement of linguistic resources during the processing of gestures rather than by the shared computational demands like communicative intent or abstract conceptual processing.

Amodal Hierarchical Processor in "Broca's Area"?

Although our action observation/imitation conditions did not include a manipulation of hierarchical complexity, we would argue that, to the extent that an amodal hierarchical processor exists in the human brain, it does not reside within the high-level language network. We have previously made this argument based on nonoverlap between language processing and music perception (e.g., Fedorenko et al. 2011, 2012c; Norman-Haignere et al. 2015). Music is another domain that has been argued to recruit such an amodal hierarchical processor (e.g., Koelsch et al. 2002; Maess et al. 2001). However, as Fedorenko and Varley (2016) have argued, the most compelling evidence for overlap comes from structure-violation paradigms, and in those paradigms, violations of structure appear to elicit similar responses to those elicited by low-level oddball manipulations (e.g., Corbetta and Shulman 2002) and plausibly arise within the domain-general MD network due to increased processing effort associated with unexpected events (Duncan 2010, 2013). Similarly, some manipulations of hierarchical complexity in the action domain (e.g., Koechlin and Jubault 2006) plausibly engage parts of the MD network because more complex action plans are associated with greater working memory and cognitive control demands. Although parts of the MD system have been argued to be particularly sensitive to hierarchical demands (e.g., Badre and D'Esposito 2007, 2009; Badre 2008) or to the level of abstractness of the to-be-processed information (e.g., Koechlin et al. 2003; Koechlin and Summerfield 2007), these proposals have not gone unchallenged (e.g., Crittenden and Duncan 2014; Pischedda et al. 2017). Thus whether an amodal hierarchical processor exists anywhere in the human brain remains an open question, but to the extent that it does, it exists outside the boundaries of the high-level language network.

Possibly Similar Computations Across Domains in Spite of Nonoverlapping Brain Regions

The existence of distinct brain regions that support information processing in different domains, like language versus action observation versus action production, does not imply that the basic computations (that operate over those domain-specific representations) are different. In fact, neural circuits across the cortex share many core properties (e.g., Douglas et al. 1989; Douglas and Martin 2004; Harris and Shepherd 2015), suggesting that the basic computations may be the same or similar across different cortical areas. It is also easy to come up with intuitive-level descriptions of potential parallels between domains. For example, in the domain of language, we have a large store of form-meaning mappings and knowledge about the relationships among them. We can use this knowledge to interpret linguistic signals, and to generate new utterances, by combining these basic building blocks into sequences. In the domain of actions, we may have a similar "vocabulary" of actions for each of our effectors associated with particular contexts of use, and information about how these actions can be combined (e.g., Hommel et al. 2001; Rosenbaum et al. 2001; Schack 2004). In addition, we can refer to this stored knowledge to interpret others' actions as well as generate our own action sequences as needed for goal-directed behavior (e.g., Flash and Bizzi 2016). As we make progress in developing fleshed-out mechanistic-level hypotheses about what actually goes on when we understand and produce language, or as we perceive and generate motor actions, it is important to keep in mind both that *1*) the linguistic and action/motor representations appear to be stored in nonoverlapping brain areas, but that *2*) the computations may be fundamentally similar between these (and possibly other domains of perception, action, and cognition).

(Ir)relevance of the Current Results to the Embodiment Debate

As noted in the INTRODUCTION, the question investigated here, i.e., whether high-level language processing brain regions are engaged when we observe or produce motor action, is distinct from the much-debated question of the nature of our conceptual representations. In particular, for many years now, some have advocated an "embodied" view of meanings whereby concepts are "grounded" in sensory-motor modalities (e.g., Barsalou et al. 2003; Simmons et al. 2007; Tranel et al. 2003). Embodiment proposals vary widely in the scope of their claims (see Leshinskaya and Caramazza 2016, for a recent review of the key issues in this debate), from a complete denial of the existence of abstract/amodal conceptual representations (e.g., Barsalou et al. 2003; Barsalou 2008; Pulvermüller and Fadiga 2010; cf. Caramazza et al. 1990) to more moderate positions where abstract representations interact in some way with the sensory/motor ones (e.g., Meteyard et al. 2012). The reason that the work reported here might, on the surface, appear to be relevant to the embodiment debate is that action verbs have received a lot of attention in that literature (e.g., see Bedny and Caramazza 2011 for a review). However, the link is superficial: whether or not sensory and/or motor brain regions are active (to some extent) when we understand the meanings of verbs like "kick" or "punch" (e.g., Hauk et al. 2004) is orthogonal to the question of whether the regions of the language network, which we know are engaged when we process word meanings (e.g., Fedorenko et al. 2012b), play a role in the processing or execution of motor actions. We here show that the answer to the latter question is no.

Conclusion

Before concluding, it is worth noting that, in general, the construct of "actions" is complex and heterogeneous, and different researchers have different notions and scope in mind when they talk about "actions." The conditions we included in our study have spanned goal-directed/transitive actions (e.g., manipulating an object in *experiment 1*) and intransitive ones (e.g., tapping a finger in *experiment 3a*), as well as actions performed by different effectors (hand, feet, eyes, mouth, and face). There are plausibly other dimensions of actions that affect their representation and processing (e.g., Tarhan and Konkle 2017): e.g., whether the action is directed toward an animate entity vs. an object, whether the action has communicative intent, whether the action maps onto a linguistic label, etc. Our study leaves open the possibility that actions that have some specific property(ies) may elicit responses in the language cortex. However, to evaluate this possibility, we need clear testable hypotheses that would formally specify the relevant features of actions that may lead to the recruitment of the same machinery as language comprehension.

To conclude, action observation and action imitation do not recruit the left-lateralized high-level language processing network, providing further evidence for the selectivity of this network for language processing (e.g., Fedorenko and Varley 2016). However, this separability is still compatible with hypotheses about the gestural origins of human language (e.g., Corballis 2003; Tomasello 2008). Furthermore, given the general similarity of neural circuits across the neocortex (e.g., Harris and Shepherd 2015), research in the domains of action perception or motor control may inform our understanding of the computations that support language comprehension and production, domains where we do not have the luxury of animal models to richly characterize neural response properties and their interactions.

ACKNOWLEDGMENTS

We acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT and its support team (Steve Shannon, Atsushi Takahashi, and Sheeba Arnold). We also thank *1*) Nancy Kanwisher for helpful discussions of the design of *experiment 1*; *2*) Anastasia (Stacey) Vishnevetsky for help in creating the materials for *experiment 1*, Jeanne Gallee for help in creating the materials for *experiment 3*, and Michael Behr for help with the early versions of the script for *experiment 1*; *3*) EvLab members for help with data collection; *4*) Matt Siegelman for help with creating the webpage for this manuscript; and *5*) Leyla Tarhan for comments on the draft of the manuscript.

GRANTS

E. Fedorenko was supported by National Institutes of Health Grants R00-HD-057522 and R01-DC-016607 and by a grant from the Simons Foundation to the Simons Center for the Social Brain at MIT.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

B.L.P., C.H., K.K., E.D., and E.F. conceived and designed research; C.H., K.K., and E.D. performed experiments; B.L.P. and C.H. analyzed data; B.L.P. and E.F. interpreted results of experiments; B.L.P. prepared figures; B.L.P. and E.F. drafted manuscript; B.L.P., K.K., E.D., and E.F. edited and revised manuscript; B.L.P., C.H., K.K., E.D., and E.F. approved final version of manuscript;

REFERENCES

Amit E, Hoeflin C, Hamzah N, Fedorenko E. An asymmetrical relationship between verbal and visual thinking: Converging evidence from behavior and fMRI. *Neuroimage* 152: 619–627, 2017. doi:10.1016/j.neuroimage.2017.03.029.

Amunts K, Lenzen M, Friederici AD, Schleicher A, Morosan P, Palomero-Gallagher N, Zilles K. Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biol* 8: e1000489, 2010. doi:10.1371/journal.pbio.1000489.

Amunts K, Schleicher A, Bürgel U, Mohlberg H, Uylings HBM, Zilles K. Broca's region revisited: cytoarchitecture and intersubject variability. *J Comp Neurol* 412: 319–341, 1999. doi:10.1002/(SICI)1096-9861(19990920)412:2<319::AID-CNE10>3.0.CO;2-7.

Andric M, Solodkin A, Buccino G, Goldin-Meadow S, Rizzolatti G, Small SL. Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia* 51: 1619–1629, 2013. doi:10.1016/j.neuropsychologia.2013.03.022.

Arbib MA. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav Brain Sci* 28: 105–124, 2005. doi:10.1017/S0140525X05000038.

Arbib MA. Mirror system activity for action and language is embedded in the integration of dorsal and ventral pathways. *Brain Lang* 112: 12–24, 2010. doi:10.1016/j.bandl.2009.10.001.

Badre D. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn Sci* 12: 193–200, 2008. doi:10.1016/j.tics.2008.02.004.

Badre D, D'Esposito M. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J Cogn Neurosci* 19: 2082–2099, 2007. doi:10.1162/jocn.2007.19.12.2082.

Badre D, D'Esposito M. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat Rev Neurosci* 10: 659–669, 2009. doi:10.1038/nrn2667.

Barsalou LW. Grounded cognition. *Annu Rev Psychol* 59: 617–645, 2008. doi:10.1146/annurev.psych.59.103006.093639.

Barsalou LW, Kyle Simmons W, Barbey AK, Wilson CD. Grounding conceptual knowledge in modality-specific systems. *Trends Cogn Sci* 7: 84–91, 2003. doi:10.1016/S1364-6613(02)00029-3.

Basilakos A, Smith KG, Fillmore P, Fridriksson J, Fedorenko E. Functional characterization of the human speech articulation network. *Cereb Cortex* 28: 1866–1830, 2018. doi:10.1093/cercor/bhx100.

Bautista A, Wilson SM. Neural responses to grammatically and lexically degraded speech. *Lang Cogn Neurosci* 31: 567–574, 2016. doi:10.1080/23273798.2015.1123281.

Bedny M, Caramazza A. Perception, action, and word meanings in the human brain: the case from action verbs. *Ann N Y Acad Sci* 1224: 81–95, 2011. doi:10.1111/j.1749-6632.2011.06013.x.

Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29: 1165–1188, 2001.

Biagi L, Cioni G, Fogassi L, Guzzetta A, Sgandurra G, Tosetti M. Action observation network in childhood: a comparative fMRI study with adults. *Dev Sci* 19: 1075–1086, 2016. doi:10.1111/desc.12353.

Binder JR, Desai RH, Graves WW, Conant LL. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19: 2767–2796, 2009. doi:10.1093/cercor/bhp055.

Binkofski F, Amunts K, Stephan KM, Posse S, Schormann T, Freund HJ, Zilles K, Seitz RJ. Broca's region subserves imagery of motion: a combined cytoarchitectonic and fMRI study. *Hum Brain Mapp* 11: 273–285, 2000. doi:10.1002/1097-0193(200012)11:4<273::AID-HBM40>3.0.CO;2-0.

Blank I, Balewski Z, Mahowald K, Fedorenko E. Syntactic processing is distributed across the language system. *Neuroimage* 127: 307–323, 2016. doi:10.1016/j.neuroimage.2015.11.069.

Blank I, Kanwisher N, Fedorenko E. A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *J Neurophysiol* 112: 1105–1118, 2014. doi:10.1152/jn.00884.2013.

Bohland JW, Guenther FH. An fMRI investigation of syllable sequence production. *Neuroimage* 32: 821–841, 2006. doi:10.1016/j.neuroimage.2006.04.173.

Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495: 327–332, 2013. *Nature* 498: 526, 2013. doi:10.1038/nature12182.

Buckner RL, Krienen FM. The evolution of distributed association networks in the human brain. *Trends Cogn Sci* 17: 648–665, 2013. doi:10.1016/j.tics.2013.09.017.

Caramazza A, Hillis AE, Rapp BC, Romani C. The multiple semantics hypothesis: multiple confusions? *Cogn Neuropsychol* 7: 161–189, 1990. doi:10.1080/02643299008253441.

Caspers S, Zilles K, Laird AR, Eickhoff SB. ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage* 50: 1148–1167, 2010. doi:10.1016/j.neuroimage.2009.12.112.

Chai LR, Mattar MG, Blank IA, Fedorenko E, Bassett DS. Functional network dynamics of the language system. *Cereb Cortex* 26: 4148–4159, 2016. doi:10.1093/cercor/bhw238.

Clerget E, Winderickx A, Fadiga L, Olivier E. Role of Broca's area in encoding sequential human actions: a virtual lesion study. *Neuroreport* 20: 1496–1499, 2009. doi:10.1097/WNR.0b013e3283329be8.

Corballis MC. *From Hand to Mouth: the Origins of Language*. Princeton, NJ: Princeton University Press, 2003.

Corballis MC. Mirror neurons and the evolution of language. *Brain Lang* 112: 25–35, 2010. doi:10.1016/j.bandl.2009.02.002.

Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3: 201–215, 2002. doi:10.1038/nrn755.

Crittenden BM, Duncan J. Task difficulty manipulation reveals multiple demand activity but no frontal lobe hierarchy. *Cereb Cortex* 24: 532–540, 2014. doi:10.1093/cercor/bhs333.

Culham JC, Valyear KF. Human parietal cortex in action. *Curr Opin Neurobiol* 16: 205–212, 2006. doi:10.1016/j.conb.2006.03.005.

D'Ausilio A, Pulvermüller F, Salmas P, Bufalari I, Begliomini C, Fadiga L. The motor somatotopy of speech perception. *Curr Biol* 19: 381–385, 2009. doi:10.1016/j.cub.2009.01.017.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31: 968–980, 2006. doi:10.1016/j.neuroimage.2006.01.021.

Dinstein I, Hasson U, Rubin N, Heeger DJ. Brain areas selective for both observed and executed movements. *J Neurophysiol* 98: 1415–1427, 2007. doi:10.1152/jn.00238.2007.

Douglas RJ, Martin KA. Neuronal circuits of the neocortex. *Annu Rev Neurosci* 27: 419–451, 2004. doi:10.1146/annurev.neuro.27.070203.144152.

Douglas RJ, Martin KA, Whitteridge D. A canonical microcircuit for neocortex. *Neural Comput* 1: 480–488, 1989. doi:10.1162/neco.1989.1.4.480.

Duncan J. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci* 14: 172–179, 2010. doi:10.1016/j.tics.2010.01.004.

Duncan J. The structure of cognition: attentional episodes in mind and brain. *Neuron* 80: 35–50, 2013. doi:10.1016/j.neuron.2013.09.015.

Duncan J, Owen AM. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci* 23: 475–483, 2000. doi:10.1016/S0166-2236(00)01633-7.

Enrici I, Adenzato M, Cappa S, Bara BG, Tettamanti M. Intention processing in communication: a common brain network for language and gestures. *J Cogn Neurosci* 23: 2415–2431, 2011. doi:10.1162/jocn.2010.21594.

Fazio P, Cantagallo A, Craighero L, D'Ausilio A, Roy AC, Pozzo T, Calzolari F, Granieri E, Fadiga L. Encoding of human action in Broca's area. *Brain* 132: 1980–1988, 2009. doi:10.1093/brain/awp118.

Fedorenko E. The role of domain-general cognitive control in language comprehension. *Front Psychol* 5: 335, 2014. doi:10.3389/fpsyg.2014.00335.

Fedorenko E, Behr MK, Kanwisher N. Functional specificity for high-level linguistic processing in the human brain. *Proc Natl Acad Sci U S A* 108: 16428–16433, 2011. doi:10.1073/pnas.1112937108.

Fedorenko E, Duncan J, Kanwisher N. Language-selective and domain-general regions lie side by side within Broca's area. *Curr Biol* 22: 2059–2062, 2012a. doi:10.1016/j.cub.2012.09.011.

Fedorenko E, Duncan J, Kanwisher N. Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci U S A* 110: 16616–16621, 2013. doi:10.1073/pnas.1315235110.

Fedorenko E, Hsieh PJ, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J Neurophysiol* 104: 1177–1194, 2010. doi:10.1152/jn.00032.2010.

Fedorenko E, McDermott JH, Norman-Haignere S, Kanwisher N. Sensitivity to musical structure in the human brain. *J Neurophysiol* 108: 3289–3300, 2012c. doi:10.1152/jn.00209.2012.

Fedorenko E, Nieto-Castañon A, Kanwisher N. Lexical and syntactic representations in the brain: an fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia* 50: 499–513, 2012b. doi:10.1016/j.neuropsychologia.2011.09.014.

Fedorenko E, Thompson-Schill SL. Reworking the language network. *Trends Cogn Sci* 18: 120–126, 2014. doi:10.1016/j.tics.2013.12.006.

Fedorenko E, Varley R. Language and thought are not the same thing: evidence from neuroimaging and neurological patients. *Ann N Y Acad Sci* 1369: 132–153, 2016. doi:10.1111/nyas.13046.

Fiebach CJ, Schubotz RI. Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains? *Cortex* 42: 499–502, 2006. doi:10.1016/S0010-9452(08)70386-1.

Fischl B, Rajendran N, Busa E, Augustinack J, Hinds O, Yeo BTT, Mohlberg H, Amunts K, Zilles K. Cortical folding patterns and predicting cytoarchitecture. *Cereb Cortex* 18: 1973–1980, 2008. doi:10.1093/cercor/bhm225.

Flash T, Bizzi E. Cortical circuits and modules in movement generation: experiments and theories. *Curr Opin Neurobiol* 41: 174–178, 2016. doi:10.1016/j.conb.2016.09.013.

Flinker A, Korzeniewska A, Shestyuk AY, Franaszczuk PJ, Dronkers NF, Knight RT, Crone NE. Redefining the role of Broca's area in speech. *Proc Natl Acad Sci U S A* 112: 2871–2875, 2015. doi:10.1073/pnas.1414491112.

Frost MA, Goebel R. Measuring structural-functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage* 59: 1369–1381, 2012. doi:10.1016/j.neuroimage.2011.08.035.

Gallivan JP, Culham JC. Neural coding within human brain areas involved in actions. *Curr Opin Neurobiol* 33: 141–149, 2015. doi:10.1016/j.conb.2015.03.012.

Goldin-Meadow S, Brentari D. Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behav Brain Sci* 40: 141–e46, 2017.

Guenther FH. *Neural Control of Speech*. Cambridge, MA: MIT Press, 2016.

Häberling IS, Corballis PM, Corballis MC. Language, gesture, and handedness: Evidence for independent lateralized networks. *Cortex* 82: 72–85, 2016. doi:10.1016/j.cortex.2016.06.003.

Harris KD, Shepherd GM. The neocortical circuit: themes and variations. *Nat Neurosci* 18: 170–181, 2015. doi:10.1038/nn.3917.

Hauk O, Johnsrude I, Pulvermüller F. Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41: 301–307, 2004. doi:10.1016/S0896-6273(03)00838-9.

Hickok G. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J Cogn Neurosci* 21: 1229–1243, 2009. doi:10.1162/jocn.2009.21189.

Hickok G, Okada K, Serences JT. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J Neurophysiol* 101: 2725–2732, 2009. doi:10.1152/jn.91099.2008.

Higuchi S, Chaminade T, Imamizu H, Kawato M. Shared neural correlates for language and tool use in Broca's area. *Neuroreport* 20: 1376–1381, 2009. doi:10.1097/WNR.0b013e3283315570.

Hommel B, Müsseler J, Aschersleben G, Prinz W. The Theory of Event Coding (TEC): a framework for perception and action planning. *Behav Brain Sci* 24: 849–878, 2001. doi:10.1017/S0140525X01000103.

Hugdahl K, Raichle ME, Mitra A, Specht K. On the existence of a generalized non-specific task-dependent network. *Front Hum Neurosci* 9: 430, 2015. doi:10.3389/fnhum.2015.00430.

Johnson-Frey SH, Newman-Norlund R, Grafton ST. A distributed left hemisphere network active during planning of everyday tool use skills. *Cereb Cortex* 15: 681–695, 2005. doi:10.1093/cercor/bhh169.

Juch H, Zimine I, Seghier ML, Lazeyras F, Fasel JH. Anatomical variability of the lateral frontal lobe surface: implication for intersubject variability in language neuroimaging. *Neuroimage* 24: 504–514, 2005. doi:10.1016/j.neuroimage.2004.08.037.

Julian JB, Fedorenko E, Webster J, Kanwisher N. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* 60: 2357–2364, 2012. doi:10.1016/j.neuroimage.2012.02.055.

Kimura D. Acquisition of a motor skill after left-hemisphere damage. *Brain* 100: 527–542, 1977. doi:10.1093/brain/100.3.527.

Kimura D, Battison R, Lubert B. Impairment of nonlinguistic hand movements in a deaf aphasic. *Brain Lang* 3: 566–571, 1976. doi:10.1016/0093-934X(76)90047-X.

Koechlin E, Jubault T. Broca's area and the hierarchical organization of human behavior. *Neuron* 50: 963–974, 2006. doi:10.1016/j.neuron.2006.05.017.

Koechlin E, Ody C, Kouneiher F. The architecture of cognitive control in the human prefrontal cortex. *Science* 302: 1181–1185, 2003. doi:10.1126/science.1088545.

Koechlin E, Summerfield C. An information theoretical approach to prefrontal executive function. *Trends Cogn Sci* 11: 229–235, 2007. doi:10.1016/j.tics.2007.04.005.

Koelsch S, Gunter TC, v Cramon DY, Zysset S, Lohmann G, Friederici AD. Bach speaks: a cortical "language-network" serves the processing of music. *Neuroimage* 17: 956–966, 2002. doi:10.1006/nimg.2002.1154.

Kriegeskorte N. Pattern-information analysis: from stimulus decoding to computational-model testing. *Neuroimage* 56: 411–421, 2011. doi:10.1016/j.neuroimage.2011.01.061.

Leshinskaya A, Caramazza A. For a cognitive neuroscience of concepts: moving beyond the grounding issue. *Psychon Bull Rev* 23: 991–1001, 2016. doi:10.3758/s13423-015-0870-z.

Lingnau A, Gesierich B, Caramazza A. Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proc Natl Acad Sci U S A* 106: 9925–9930, 2009. doi:10.1073/pnas.0902262106.

Maess B, Koelsch S, Gunter TC, Friederici AD. Musical syntax is processed in Broca's area: an MEG study. *Nat Neurosci* 4: 540–545, 2001. doi:10.1038/87502.

Mahowald K, Fedorenko E. Reliable individual-level neural markers of high-level language processing: a necessary precursor for relating neural variability to behavioral and genetic variability. *Neuroimage* 139: 74–93, 2016. doi:10.1016/j.neuroimage.2016.05.073.

Marstaller L, Burianová H. The multisensory perception of co-speech gestures – A review and meta-analysis of neuroimaging studies. *J Neurolinguist* 30: 69–77, 2014. doi:10.1016/j.jneuroling.2014.04.003.

McNeill D. *Hand and Mind: What Gestures Reveal About Thought*. Chicago, OH: University of Chicago Press, 1992.

Meister IG, Iacoboni M. No language-specific activation during linguistic processing of observed actions. *PLoS One* 2: e891, 2007. doi:10.1371/journal.pone.0000891.

Menenti L, Gierhan SME, Segaert K, Hagoort P. Shared language: overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychol Sci* 22: 1173–1182, 2011. doi:10.1177/0956797611418347.

Meteyard L, Cuadrado SR, Bahrami B, Vigliocco G. Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex* 48: 788–804, 2012. doi:10.1016/j.cortex.2010.11.002.

Molenberghs P, Cunnington R, Mattingley JB. Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. *Neurosci Biobehav Rev* 36: 341–349, 2012. doi:10.1016/j.neubiorev.2011.07.004.

Monti MM, Osherson DN. Logic, language and the brain. *Brain Res* 1428: 33–42, 2012. doi:10.1016/j.brainres.2011.05.061.

Mukamel R, Ekstrom AD, Kaplan J, Iacoboni M, Fried I. Single-neuron responses in humans during execution and observation of actions. *Curr Biol* 20: 750–756, 2010. doi:10.1016/j.cub.2010.02.045.

Murakami T, Restle J, Ziemann U. Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. *Neuropsychologia* 49: 2045–2054, 2011. doi:10.1016/j.neuropsychologia.2011.03.034.

Nieto-Castañón A, Fedorenko E. Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage* 63: 1646–1669, 2012. doi:10.1016/j.neuroimage.2012.06.065.

Norman-Haignere S, Kanwisher NG, McDermott JH. Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88: 1281–1296, 2015. doi:10.1016/j.neuron.2015.11.035.

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9: 97–113, 1971. doi:10.1016/0028-3932(71)90067-4.

Overath T, McDermott JH, Zarate JM, Poeppel D. The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nat Neurosci* 18: 903–911, 2015. doi:10.1038/nn.4021.

Papagno C, Della Sala S, Basso A. Ideomotor apraxia without aphasia and aphasia without apraxia: the anatomical support for a double dissociation. *J Neurol Neurosurg Psychiatry* 56: 286–289, 1993. doi:10.1136/jnnp.56.3.286.

Petrides M, Pandya DN. Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol* 7: e1000170, 2009. doi:10.1371/journal.pbio.1000170.

Pierrot-Deseilligny C, Milea D, Müri RM. Eye movement control by the cerebral cortex. *Curr Opin Neurol* 17: 17–25, 2004. doi:10.1097/00019052-200402000-00005.

Pischedda D, Görgen K, Haynes JD, Reverberi C. neural representations of hierarchical rule sets: the human control system represents rules irrespective of the hierarchical level to which they belong. *J Neurosci* 37: 12281–12296, 2017. doi:10.1523/JNEUROSCI.3088-16.2017.

Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N. Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage* 56: 2356–2363, 2011. doi:10.1016/j.neuroimage.2011.03.067.

Poldrack RA. Can cognitive processes be inferred from neuroimaging data? *Trends Cogn Sci* 10: 59–63, 2006. doi:10.1016/j.tics.2005.12.004.

Poldrack RA. Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron* 72: 692–697, 2011. doi:10.1016/j.neuron.2011.11.001.

Pulvermüller F, Fadiga L. Active perception: sensorimotor circuits as a cortical basis for language. *Nat Rev Neurosci* 11: 351–360, 2010. doi:10.1038/nrn2811.

Pulvermüller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y. Motor cortex maps articulatory features of speech sounds. *Proc Natl Acad Sci U S A* 103: 7865–7870, 2006. doi:10.1073/pnas.0509989103.

Ralph MA, Jefferies E, Patterson K, Rogers TT. The neural and computational bases of semantic cognition. *Nat Rev Neurosci* 18: 42–55, 2017. doi:10.1038/nrn.2016.150.

Redcay E, Ludlum RS, Velnoskey KR, Kanwal S. Communicative signals promote object recognition memory and modulate the right posterior STS. *J Cogn Neurosci* 28: 8–19, 2016. doi:10.1162/jocn_a_00875.

Rizzolatti G, Arbib MA. Language within our grasp. *Trends Neurosci* 21: 188–194, 1998. doi:10.1016/S0166-2236(98)01260-0.

Rizzolatti G, Camarda R, Fogassi L, Gentilucci M, Luppino G, Matelli M. Function organization of inferior area 6 in the macaque monkey. *Exp Brain Res* 71: 491–507, 1988. doi:10.1007/BF00248742.

Rosenbaum DA, Meulenbroek RJ, Vaughan J, Jansen C. Posture-based motion planning: applications to grasping. *Psychol Rev* 108: 709–734, 2001. doi:10.1037/0033-295X.108.4.709.

Saxe R, Brett M, Kanwisher N. Divide and conquer: a defense of functional localizers. *Neuroimage* 30: 1088–1096, 2006. doi:10.1016/j.neuroimage.2005.12.062.

Saygin AP, Wilson SM, Dronkers NF, Bates E. Action comprehension in aphasia: linguistic and non-linguistic deficits and their lesion correlates. *Neuropsychologia* 42: 1788–1804, 2004. doi:10.1016/j.neuropsychologia.2004.04.016.

Schack T. The cognitive architecture of complex movement. *Int J Sport Exerc Psychol* 2: 403–438, 2004. doi:10.1080/1612197X.2004.9671753.

Scott TL, Gallée J, Fedorenko E. A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cogn Neurosci* 8: 167–176, 2017.

Shepherd SV, Freiwald WA. Functional networks for social communication in the macaque monkey. *Neuron* 99: 413–420.e3, 2018. doi:10.1016/j.neuron.2018.06.027.

Simmons WK, Ramjee V, Beauchamp MS, McRae K, Martin A, Barsalou LW. A common neural substrate for perceiving and knowing about color. *Neuropsychologia* 45: 2802–2810, 2007. doi:10.1016/j.neuropsychologia.2007.05.002.

Sirigu A, Cohen L, Zalla T, Pradat-Diehl P, Van Eeckhout P, Grafman J, Agid Y. Distinct frontal regions for processing sentence syntax and story grammar. *Cortex* 34: 771–778, 1998. doi:10.1016/S0010-9452(08)70780-9.

Slocombe KE. Vocal communication in primates In: *Emerging Trends in the Social and Behavioral Sciences: an Interdisciplinary, Searchable, and Linkable Resource*, edited by Kosslyn SM. Hoboken, NJ: Wiley, 2015, p. 1–12. doi:10.1002/9781118900772.etrds0380

Tahmasebi AM, Davis MH, Wild CJ, Rodd JM, Hakyemez H, Abolmaesumi P, Johnsrude IS. Is the link between anatomical structure and function equally strong at all cognitive levels of processing? *Cereb Cortex* 22: 1593–1603, 2012. doi:10.1093/cercor/bhr205.

Tarhan L, Konkle T. Modeling the neural structure underlying human action perception (Online). *Conference on Cognitive Computational Neuroscience*. https://www2.securecms.com/CCNeuro/docs-0/591b2fa768ed3fa74b57b899.pdf [2017].

Tettamanti M, Weniger D. Broca's area: a supramodal hierarchical processor? *Cortex* 42: 491–494, 2006. doi:10.1016/S0010-9452(08)70384-8.

Thesen S, Heid O, Mueller E, Schad LR. Prospective acquisition correction for head motion with image-based tracking for real-time fMRI. *Magn Reson Med* 44: 457–465, 2000. doi:10.1002/1522-2594(200009)44:3<457::AID-MRM17>3.0.CO;2-R.

Tomaiuolo F, MacDonald JD, Caramanos Z, Posner G, Chiavaras M, Evans AC, Petrides M. Morphology, morphometry and probability mapping of the pars opercularis of the inferior frontal gyrus: an in vivo MRI analysis. *Eur J Neurosci* 11: 3033–3046, 1999. doi:10.1046/j.1460-9568.1999.00718.x.

Tomasello M. *Origins of Human Communication*. Cambridge, MA: MIT Press, 2008.

Tranel D, Kemmerer D, Adolphs R, Damasio H, Damasio AR. Neural correlates of conceptual knowledge for actions. *Cogn Neuropsychol* 20: 409–432, 2003. doi:10.1080/02643290244000248.

Tremblay P, Dick AS. Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain Lang* 162: 60–71, 2016. doi:10.1016/j.bandl.2016.08.004.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15: 273–289, 2002. doi:10.1006/nimg.2001.0978.

Villarreal M, Fridman EA, Amengual A, Falasco G, Gerschcovich ER, Ulloa ER, Leiguarda RC. The neural substrate of gesture recognition. *Neuropsychologia* 46: 2371–2382, 2008. doi:10.1016/j.neuropsychologia.2008.03.004.

Watkins KE, Strafella AP, Paus T. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41: 989–994, 2003. doi:10.1016/S0028-3932(02)00316-0.

Willems RM, Hagoort P. Neural evidence for the interplay between language, gesture, and action: a review. *Brain Lang* 101: 278–289, 2007. doi:10.1016/j.bandl.2007.03.004.

Willems RM, Van der Haegen L, Fisher SE, Francks C. On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nat Rev Neurosci* 15: 193–201, 2014. doi:10.1038/nrn3679.

Xu J, Gannon PJ, Emmorey K, Smith JF, Braun AR. Symbolic gestures and spoken language are processed by a common neural system. *Proc Natl Acad Sci U S A* 106: 20664–20669, 2009. doi:10.1073/pnas.0909197106.

Yang J, Andric M, Mathew MM. The neural basis of hand gesture comprehension: A meta-analysis of functional magnetic resonance imaging studies. *Neurosci Biobehav Rev* 57: 88–104, 2015. doi:10.1016/j.neubiorev.2015.08.006.