Report

Congruent Embodied Representations for Visually Presented Actions and Linguistic Phrases Describing Actions

Lisa Aziz-Zadeh,[1,2,3,8,*] Stephen M. Wilson,[4,5]
Giacomo Rizzolatti,[1] and Marco Iacoboni[4,6,7]
[1] Dipartimento di Neuroscienze
Università di Parma
Via Volturno, 39/E
I-43100 Parma
Italy
[2] Department of Psychology
University of California, Berkeley
Berkeley, California 94720
[3] International Computer Science Institute
1947 Center Street
Berkeley, California 94704
[4] Ahmanson Lovelace Brain Mapping Center
[5] Neuroscience Interdepartmental Program
[6] Department of Psychiatry and Biobehavioral Sciences
[7] Brain Research Institute
    David Geffen School of Medicine
University of California, Los Angeles
Los Angeles, California 90095

Summary

The thesis of embodied semantics holds that conceptual representations accessed during linguistic processing are, in part, equivalent to the sensory-motor representations required for the enactment of the concepts described [1–6]. Here, using fMRI, we tested the hypothesis that areas in human premotor cortex that respond both to the execution and observation of actions—mirror neuron areas [7–18]—are key neural structures in these processes. Participants observed actions and read phrases relating to foot, hand, or mouth actions. In the premotor cortex of the left hemisphere, a clear congruence was found between effector-specific activations of visually presented actions and of actions described by literal phrases. These results suggest a key role of mirror neuron areas in the re-enactment of sensory-motor representations during conceptual processing of actions invoked by linguistic stimuli.

Results

Subjects (n = 12) were scanned with fMRI while viewing actions performed by the mouth, the hand, or the foot, and reading literal and metaphorical sentences relating to the mouth, the hand, or the foot. Areas activated during action observation (averaged across the three effectors) as compared to the resting baseline, and reading (averaged across literal and metaphorical sentences and across the three effectors) as compared to rest,

*Correspondence: lisa.azizzadeh@usc.edu
[8] Present address: Brain and Creativity Institute and Department of Occupational Sciences, University of Southern California, Los Angeles, California 90089.

are depicted in Figure 1 and Table S1 in the Supplemental Data available online. Both tasks activated extensive subcortical and cortical visual areas. For action observation, these activations extended well into the dorsal visual stream and there were also bilateral activations in ventral premotor cortex. For language, there was a large area of activation in the left inferior frontal gyrus (IFG) (pars opercularis, triangularis, and orbitalis) extending into premotor cortex in the precentral gyrus, and a smaller activation in the right IFG centered in the pars opercularis.

Next we examined areas activated by observation of each of the three effectors (Figure 2A and Table S2). Observation of all three effectors was associated with bilateral activations in ventral premotor cortex. A more dorsal sector of ventral premotor cortex was activated (marginally) only by observation of hand actions (p = 0.069). Thus, within the ventral premotor cortex, responses strongest for the mouth were located ventrally, whereas responses strongest for the hand were located more dorsally. This relative arrangement, with hand dorsal to mouth, is consistent with previously described action observation maps [15, 19].

Because in this last analysis we compared activation of each effector against rest, its significance could have been a reflection of less noise in one condition as compared to the others, rather than an actual signal increase. To control for this possibility, a region-of-interest (ROI) analysis was carried out to examine signal change in the two potentially effector-specific regions in left ventral premotor cortex (Figure 2B). As expected, the more dorsal region responded most strongly to observation of the hand, and the more ventral region responded most strongly to observation of the mouth. The interaction between region and observed effector was significant [F(1, 11) = 40.72, p < 0.0001]. Note that both regions do in fact respond above baseline to observation of actions with all effectors, including the foot.

To identify possible congruence between premotor activations for action observation and phrase reading for each effector, we carried out an individual subjects analysis. For each subject, the most responsive premotor voxels in each of the three observation conditions were identified. Then we plotted the spatial locations of these peak voxels (Figure 3) and examined linguistic responses in observation-defined peak voxels.

Responses to reading of literal phrases, as well as action observation, in voxels defined by peak responses to action observation (VOI) are depicted in Figure 4. Crucially, for literal phrases, there was a significant interaction of VOI by phrase effector [F(4, 8) = 5.20, p = 0.023] in the left hemisphere (Figure 4A). Each of the three pairwise interactions was significant or marginally significant (foot versus hand: p = 0.013; foot versus mouth: p = 0.081; hand versus mouth: p = 0.061; all two-tailed). This indicates that each VOI responded most to phrases relating to the effector for which it was defined based on action observation, and constitutes evidence for
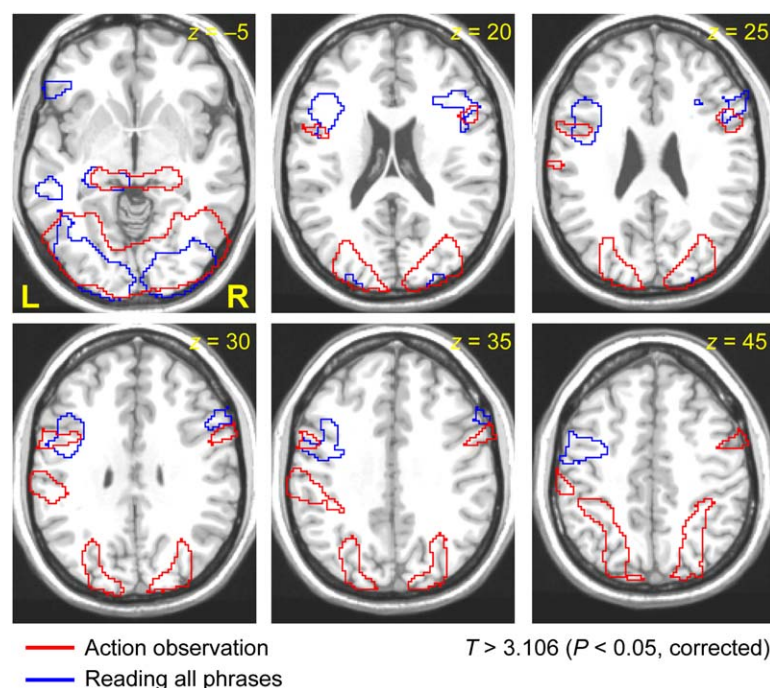
Figure 1. Activation for Observing All Actions and Reading All Phrases

Both contrasts are relative to a resting baseline.

Action observation
Reading all phrases

$T > 3.106$ ($P < 0.05$, corrected)

congruence between representations of visually presented actions and semantic representations of actions derived from linguistic phrases.

In the right hemisphere, the interaction of VOI by phrase effector was not significant [$F(4, 8) = 1.24$, $p = 0.37$] (Figure 4B). Several researchers have proposed that metaphorical sentences may rely upon embodied representations [1, 4]. For metaphorical sentences, the interaction was not significant in either the left [$F(4, 8) = 0.15$, $p = 0.96$] or the right [$F(4, 8) = 0.95$, $p = 0.48$] hemisphere. However, the negative results regarding metaphorical sentences should be interpreted with caution due to methodological considerations (see Figure S1).

Not surprisingly, voxels defined by action observation were most responsive during observation of their particular effector in both the left [interaction of VOI by

Observe mouth
Observe hand
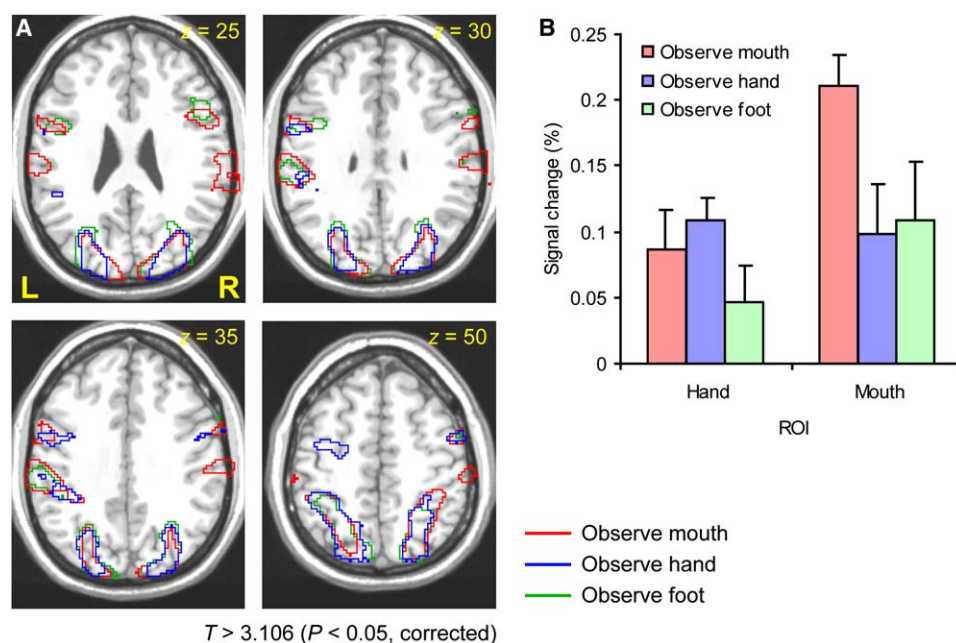Observe foot

$T > 3.106$ ($P < 0.05$, corrected)

Figure 2. Areas Activated by Observation of Mouth, Hand, and Foot Actions, and Signal Change in the Effector-Specific Regions Identified

(A) Areas activated by observation of mouth, hand, and foot actions, relative to a resting baseline.

(B) Signal change in the effector-specific regions identified in (A). Note that the mouth area responds most strongly to observation of the mouth, and the hand area most strongly to observation of the hand. Error bars represent the SEM.
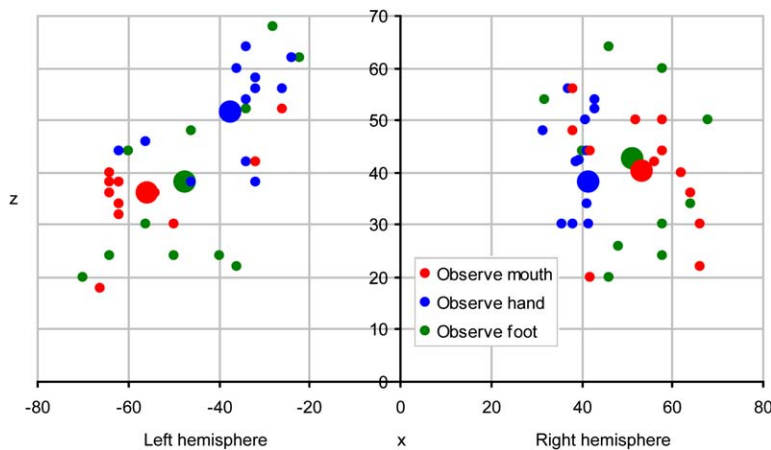
Figure 3. Coronal View Showing Premotor Peaks for Each Individual Subject for Observation of Mouth, Hand, and Foot Actions

The larger symbols show the mean locations across subjects in each hemisphere.

effector: F(4, 8) = 35.93, p < 0.0001] (Figure 4C) and right [F(4, 8) = 17.93, p = 0.0005] (Figure 4D) hemispheres. All relevant pairwise comparisons were significant. This confirms that although these peaks were identified based on contrasts against rest, each set of peaks is in fact most responsive to the effector on the basis of which it was identified.

Finally, we compared areas activated by reading literal and metaphorical phrases. The pars orbitalis of the IFG was significantly more activated for metaphorical sentences (peak: −48, 18, −16; cluster p = 0.0054); the only other significantly activated region in this contrast was a visual area close to visual motion area MT (peak = 44, −82, 16; cluster p < 0.0001). The relatively
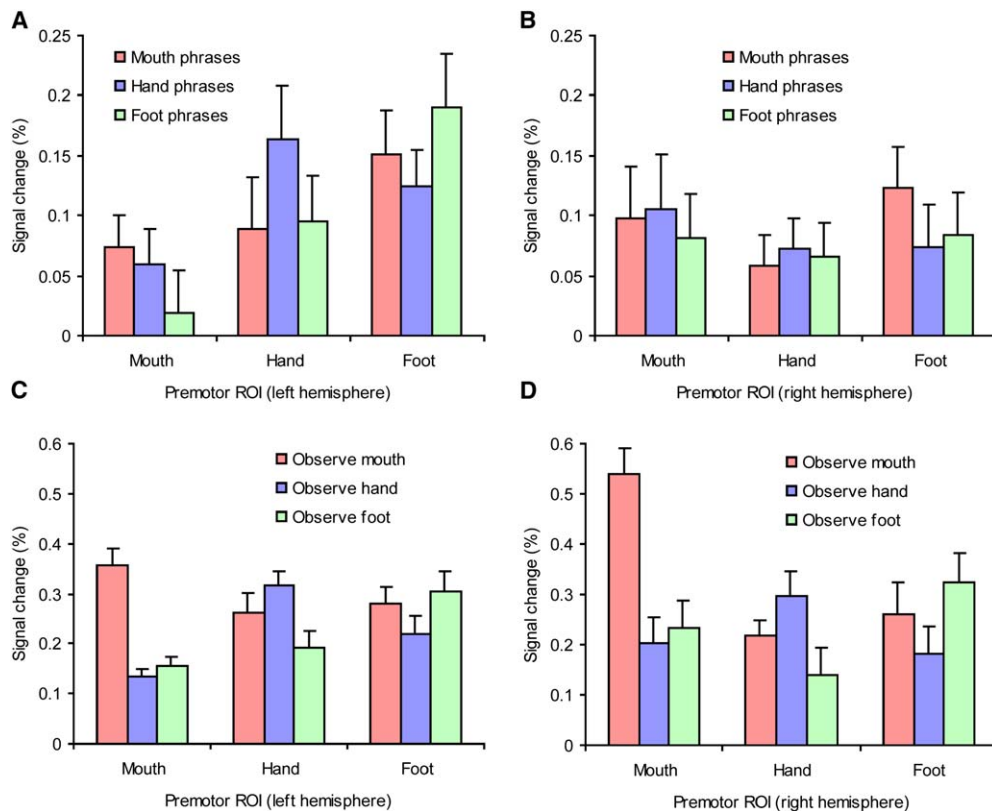

Figure 4. Signal Change for Reading Literal Phrases in ROIs Defined by Action Observation

(A) Signal change in the left premotor cortex for reading literal phrases concerning each effector, in peak voxels for action observation defined individually for each subject. Note that the interaction of voxel by effector was significant: these peak voxels show the same effector specificity for linguistic phrases that they do for action observation.

(B) As for (A) but for the right hemisphere. There was no significant interaction.

(C) Signal change in the left premotor cortex for action observation. Note that as expected, each set of peak voxels responds most strongly to the effector on the basis of which it was identified.

(D) As for (C) but in the right hemisphere.

All error bars in this figure represent SEM.

greater activation in the left pars orbitalis of the IFG (BA 47) is consistent with other recent fMRI studies [20, 21]. This portion of the IFG is thought to be involved with semantic processing [22, 23] and may thus play an important role in metaphor comprehension.

Discussion

The main aim of the present study was to determine whether phrases describing actions made with different effectors (hand, mouth, and leg) would activate those sectors of the agranular frontal cortex (motor and premotor areas) that are active when an individual observes actions made by others with the same effectors. Congruence between the cortical sectors activated by observing actions and by their verbal descriptions provides evidence for an involvement of premotor areas with mirror neuron properties in re-enactment of sensory-motor representations during conceptual processing of linguistic phrases describing actions.

These results are consistent with two previous studies reporting separate representations of linguistic stimuli differing for their action content in premotor cortex [24, 25]. These studies, however, did not compare activity related to linguistic stimuli with activity related to action observation. Nor have linguistic action representations been investigated at a subject-by-subject level.

The direct comparison of premotor representations of action concepts derived through observation and language is important because it suggests that language makes use of the same embodied representations that are thought to be involved in social cognition of the actions and intentions of others [26, 27]. The communication of concepts from the mind of one speaker to the mind of another could be accomplished with great immediacy if language utilizes the same neural representations of a concept that would be activated by direct experience of the same concept.

When investigating shared activations for action observation and action sentences in premotor cortex, it is necessary to consider the possible intermediation of motor imagery in facilitating shared activations. However, it is likely that the current data reflect direct activation of action representations in the premotor cortex during the reading of action phrases, rather than indirect activation due to the production of motor imagery. First, congruent maps for action observation and language were observed only in the left hemisphere. If these results were produced by motor imagery, congruent maps should have occurred in both hemispheres. In fact, motor imagery of the mouth and foot movements are known to activate the premotor cortex bilaterally [28–30]. Second, participants were given the language task prior to the action observation task. Although this obviously does not preclude motor imagery, it does avoid having participants read phrases relating to actions they have just observed, which might facilitate imagery.

Similarly, it is unlikely that our effects are due to covert speech during action observation. During action observation, premotor areas are activated bilaterally. Should the premotor activation during action observation been simply an effect of covert speech, we would have expected predominantly left-lateralized premotor activation as was observed in the reading conditions. Previous research has also argued against a role for covert speech in explaining activations in premotor cortex during action observation or during language comprehension [31].

In sum, these results support a key role of premotor areas with mirror neuron properties for embodied semantic representations of actions, whether they are derived through visual or linguistic modalities.

Participants

Twelve healthy right-handed volunteers (four men, eight women; mean age = 24; range = 20–37) participated in the study. All participants gave informed consent and the study was approved by the UCLA institutional review board. Handedness was determined by a modified Oldfield handedness questionnaire [32]. All participants had normal or corrected-to-normal vision. Prior to scanning, participants completed a screening questionnaire to exclude participants on medication, with a history of neurological or psychiatric disorders, substance abuse, and other medical conditions.

Stimuli and Design

There were three functional runs: linguistic phrases were presented in the first two runs, and videos of actions in the third run. This order was maintained across subjects in order to minimize motor imagery while reading the action phrases. Each run was 6 min and 32 s long and began with 8 s of rest followed by 12 blocks of 20 s. Each block was followed by 12 s of rest in which a gray screen was presented. Participants were instructed simply to read the phrases or to watch the videos. After the scan, they were asked to recall the phrases and videos presented. All participants were able to adequately report the presented stimuli. The order of the blocks in each run was random and varied across participants.

There were six kinds of language blocks obtained by crossing three effectors (mouth, hand, and foot) by two sentence types (literal and metaphorical). Examples of the phrases included: mouth/literal (e.g., "biting the peach," "biting the banana"); hand/literal (e.g., "grasping the scissors," "grasping the pen"); foot/literal (e.g., "pressing the piano pedal," "pressing the car brake"); mouth/metaphorical (e.g., "biting off more than you can chew," "chewing over the details"); hand/metaphorical (e.g., "grasping the idea," "handling the truth"); and foot/metaphorical (e.g., "kicking off the year," "time is running"). There were five sentences in each condition. Each block consisted of ten sentences (thus the five phrases were repeated twice each) displayed for 1.5 s each, with 0.5 s between phrases. Therefore, the total data for each condition consisted of 40 sentences (five sentences, which were read eight times each: two repetitions per block, by two blocks per run, by two runs).

In the action observation run, there were three kinds of blocks containing videos of mouth, hand, or foot actions. There were five videos in each condition. The mouth videos showed a side-on view of the mouth of an actor (including nose and chin) biting five different fruits (e.g., peach, banana). The hand videos showed a hand reaching for and grasping five different objects (e.g., scissors, pen). The foot videos showed the foot (and lower leg) of an actor pressing on various objects (e.g., piano pedal, car brake). Each video lasted for 1.5 s and there were 0.5 s between videos.

Image Acquisition

Images were acquired using a Siemens Allegra 3 T MRI scanner. Each of the three functional runs (TR = 2.0 s; TE = 25 ms; flip angle = 90°; 36 axial slices with interleaved acquisition; 3 × 3 × 4 mm resolution; 1 mm gap; field of view = 192 × 192 × 144 mm) comprised 196 volumes, of which the first two were discarded to allow the signal to reach steady state.

Before the functional runs, T2-weighted coplanar images were acquired (TR = 5000; TE = 33; flip angle = 90°; 36 axial slices; 1.5 × 1.5 × 4 mm resolution; 1 mm gap), and after the functional run, we acquired an MPRAGE sequence (TR = 2300, TE = 2.93, flip angle = 8°; 160 sagittal slices; 1.33 × 1.33 × 1.5 mm resolution).

Image Analysis

The fMRI data were preprocessed using tools from FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Skull stripping was performed with BET [33]. Motion correction was carried out with MCFLIRT [34, 35]. The data were smoothed with a Gaussian kernel (8 mm FWHM) and mean signal intensity was normalized across subjects using the program IP.

Statistical analysis was performed with the FMRISTAT toolbox [36]. There was a separate explanatory variable for each of the six (runs 1 and 2) or three (run 3) conditions. These design matrices were convolved with a hemodynamic response function modeled as a difference of two γ functions. Temporal drift was removed by adding a cubic spline in the frame times to the design matrix (one covariate per 2 min of scan time), and spatial drift was removed by adding a covariate in the whole volume average. Autocorrelation parameters were estimated at each voxel and used to whiten the data and design matrix. The first two functional runs, which contained the same types of blocks, were combined with a fixed effects model.

Registration was performed with the FSL tool FLIRT. For each subject, functional images were aligned to high-resolution coplanar images using an affine transformation with six degrees of freedom. High-resolution coplanar images were in turn aligned to MPRAGE images, with an affine transformation with six degrees of freedom. Finally, MPRAGE images were aligned to the standard MNI average of 152 brains using an affine transformation with 12 degrees of freedom.

Group Analysis

Group analysis was performed with FMRISTAT with a mixed effects (also known as random effects) linear model. Standard deviations from individual subject analyses were passed up to the group level. The resulting t statistic images were thresholded at t > 3.106 (df = 11, p < 0.005 uncorrected) at the voxel level, with a minimum cluster size then applied so that only clusters significant at p < 0.05 (corrected based on Gaussian random field theory) were reported. For action observation of each particular effector, reduced minimum cluster sizes were required for the cluster which was closest to the a priori region of interest of the posterior inferior frontal gyrus (pars opercularis), according to the method described in [37]. In order to display multiple activation maps simultaneously, we used a custom MATLAB program to display outlines of significant clusters on a high-resolution single subject T1 image [38].

Individual Subject Analysis

The region-of-interest (ROI) analysis in Figure 2B was carried out using MATLAB by defining ROIs based on premotor areas responsive to action observation of the mouth or hand at the threshold described above and then determining mean percent signal change across the ROI for each effector.

In the final analysis, we found the most responsive premotor voxels in each hemisphere of each subject for each of the three action observation conditions (foot, hand, mouth). For the purpose of identifying "premotor" voxels, we defined premotor cortex functionally as the region activated at low threshold (t > 1.796, p < 0.05 uncorrected) by observation of all actions versus rest in the group data, dilated by 12 mm. This area roughly encompassed the precentral gyrus and the posterior parts of the inferior and middle frontal gyri. We then plotted the mean locations of these peak voxels across subjects (Figure 3) and used repeated-measures ANOVAs to compare responses to action observation and linguistic phrases in these voxels to determine whether signal change was dependent on effector (Figure 4).

Supplemental Data

Supplemental Data include two tables and one figure and are available with this article online at http://www.current-biology.com/cgi/content/full/16/18/1818/DC1/.

References

1. Gallese, V., and Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in reason and language. Cogn. Neuropsychol. 22, 455–479.

2. Glenberg, A.M. (1997). What memory is for. Behav. Brain Sci. 20, 1–19.

3. Feldman, J., and Narayanan, S. (2004). Embodied meaning in a neural theory of language. Brain Lang. 89, 385–392.

4. Lakoff, G., and Johnson, M. (1999). Philosophy in the flesh: The embodied mind and its challenge to western thought (New York: Basic Books).

5. Barsalou, L.W. (1999). Perceptual symbol systems. Behav. Brain Sci. 22, 577–609.

6. Glenberg, A.M., and Kaschak, M.P. (2002). Grounding language in action. Psychon. Bull. Rev. 9, 558–565.

7. Johnson-Frey, S.H., Maloof, F.R., Newman-Norlund, R., Farrer, C., Inati, S., and Grafton, S.T. (2003). Actions or hand-object interactions? Human inferior frontal cortex and action observation. Neuron 39, 1053–1058.

8. Grezes, J., Costes, N., and Decety, J. (1998). Top-down effect of strategy on the perception of human biological motion: A PET investigation. Cogn. Neuropsychol. 15, 553–582.

9. Grafton, S.T., Arbib, M.A., Fadiga, L., and Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. Exp. Brain Res. 112, 103–111.

10. Nishitani, N., and Hari, R. (2000). Temporal dynamics of cortical representation for action. Proc. Natl. Acad. Sci. USA 97, 913–918.

11. Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. Annu. Rev. Neurosci. 27, 169–192.

12. Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. Science 308, 662–667.

13. Aziz-Zadeh, L., Koski, L., Zaidel, E., Mazziotta, J., and Iacoboni, M. (2006). Lateralization of the human mirror neuron system. J. Neurosci. 26, 2964–2970.

14. Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C., and Rizzolatti, G. (1999). Cortical mechanisms of human imitation. Science 286, 2526–2528.

15. Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., and Freund, H.J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. Eur. J. Neurosci. 13, 400–404.

16. Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., and Fazio, F. (1996). Localization of grasp representations in humans by PET: 1. Observation versus execution. Exp. Brain Res. 111, 246–252.

17. Koski, L., Iacoboni, M., Dubeau, M.C., Woods, R.P., and Mazziotta, J.C. (2003). Modulation of cortical activity during different imitative behaviors. J. Neurophysiol. 89, 460–471.

18. Grezes, J., Armony, J.L., Rowe, J., and Passingham, R.E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: An fMRI study. Neuroimage 18, 928–937.

19. Wheaton, K.J., Thompson, J.C., Syngeniotis, A., Abbott, D.F., and Puce, A. (2004). Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. Neuroimage *22*, 277–288.

20. Rapp, A.M., Leube, D.T., Erb, M., Grodd, W., and Kircher, T.T. (2004). Neural correlates of metaphor processing. Brain Res. Cogn. Brain Res. *20*, 395–402.

21. Stringaris, A.K., Medford, N.C., Giampietro, V., Brammer, M.J., and David, A.S. (2005). Deriving meaning: Distinct neural mechanisms for metaphoric, literal, and non-meaningful sentences. Brain Lang., in press. Published online September 13, 2005. 10.1016/j.bandl.2005.08.001.

22. Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. Annu. Rev. Neurosci. *25*, 151–188.

23. Dapretto, M., and Bookheimer, S.Y. (1999). Form and content: Dissociating syntax and semantics in sentence comprehension. Neuron *24*, 427–432.

24. Hauk, O., Johnsrude, I., and Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. Neuron *41*, 301–307.

25. Tettamanti, M., Buccino, G., Saccuman, M.C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S.F., and Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. J. Cogn. Neurosci. *17*, 273–281.

26. Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J.C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. PLoS Biol. *3*, e79. 10.1371/journal.pbio.0030079.

27. Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. Trends Cogn. Sci. *8*, 396–403.

28. Ehrsson, H.H., Geyer, S., and Naito, E. (2003). Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations. J. Neurophysiol. *90*, 3304–3316.

29. Hanakawa, T., Honda, M., Sawamoto, N., Okada, T., Yonekura, Y., Fukuyama, H., and Shibasaki, H. (2002). The role of rostral Brodmann area 6 in mental-operation tasks: An integrative neuroimaging approach. Cereb. Cortex *12*, 1157–1170.

30. Johnson, S.H., Rotte, M., Grafton, S.T., Hinrichs, H., Gazzaniga, M.S., and Heinze, H.J. (2002). Selective activation of a parieto-frontal circuit during implicitly imagined prehension. Neuroimage *17*, 1693–1704.

31. Hauk, O., and Pulvermuller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. Hum. Brain Mapp. *21*, 191–201.

32. Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia *9*, 97–113.

33. Smith, S. (2002). Fast robust automated brain extraction. Hum. Brain Mapp. *17*, 143–155.

34. Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. Med. Image Anal. *5*, 143–156.

35. Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. Neuroimage *17*, 825–841.

36. Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., and Evans, A.C. (2002). A general statistical analysis for fMRI data. Neuroimage *15*, 1–15.

37. Friston, K.J. (1997). Testing for anatomically specified regional effects. Hum. Brain Mapp. *5*, 133–136.

38. Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., and Evans, A.C. (1998). Enhancement of MR images using registration for signal averaging. J. Comput. Assist. Tomogr. *22*, 324–333.