# Motion and actions in language: Semantic representations in occipito-temporal cortex

Gina F. Humphreys, Katherine Newling, Caroline Jennings, Silvia P. Gennari *

Department of Psychology, University of York, Heslington YO10 5DD, United Kingdom

## ARTICLE INFO

## ABSTRACT

Understanding verbs typically activates posterior temporal regions and, in some circumstances, motion perception area V5. However, the nature and role of this activation remains unclear: does language alone indeed activate V5? And are posterior temporal representations modality-specific motion representations, or supra-modal motion-independent event representations? Here, we address these issues by investigating human and object motion sentences compared to corresponding state descriptions. We adopted the blank screen paradigm, which is known to encourage visual imagery, and used a localizer to identify V5 and temporal structures responding to motion. Analyses in each individual brain suggested that language modulated activity in the posterior temporal lobe but not within V5 in most participants. Moreover, posterior temporal structures strongly responded to both motion sentences and human static sentences. These results suggest that descriptive language alone need not recruit V5 and instead engages more schematic event representations in temporal cortex encoding animacy and motion.

© 2013 Published by Elsevier Inc.

## 1. Introduction

Language is often used to describe events in the world. In this function, words and sentences elicit mental representations of objects and events that have been systematically linked to aspects of experience during the course of learning. What is the nature of these representations? Are they similar to pictures or mental images of the world that are closely linked to the experiences from which they were formed? Or are they schematic and somewhat removed from that experience? And if they are schematic in nature, what sort of semantic information do they encode? These questions have been extensively investigated in the field of embodied cognition, particularly regarding motor-related properties of words. Various approaches have been identified, which provide a continuum of theoretical positions varying in two main dimensions investigated here: the degree of sensory-motor specificity attributed to linguistic representations (e.g., whether they are modality-specific, supramodal, or amodal) and the degree of interdependence proposed between linguistic and sensory-motor representations (Binder & Desai, 2011; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012).

In this paper, we explore these aspects of semantic representations with specific reference to descriptions of motion events. At least three views have been put forward in the motion literature.

One view, referred to as *weak embodiment,* suggests that linguistic and conceptual action representations are organized around and distributed along sensory cortices, but linguistic representations need not be identical to perceptual ones (Barsalou, Simmons, Barbey, & Wilson, 2003; Kable, Lease-Spellmeyer, & Chatterjee, 2002; Martin & Chao, 2001). Motion verbs for example, engage posterior temporal brain regions that partially overlap with those elicited by visual perception (Bedny, Caramazza, Grossman, Pascual-Leone, & Saxe, 2008; Kable et al., 2002) but critically, these regions need not include area V5, an area well known to selectively respond to visual motion perception (Beauchamp, Lee, Haxby, & Martin, 2002; Dumoulin et al., 2000; Tootell et al., 1995; Zihl, von Cramon, Mai, & Schmid, 1991). This kind of findings has led to the proposal of a specificity gradient along occipito-temporal structures going from more posterior modality-specific perceptual motion representations to more anterior, integrative and schematic motion event representations responding to various modalities (perceptual motion, motion language, auditory motion (Beauchamp, 2005)). This arrangement suggests a partial dependency and partial overlap between linguistic and sensory-motor representations: linguistic representations of motion are abstracted or schematized from experience during development and are thus encoded near those sensory regions processing relevant input, although these representations are supra-modal in nature.

The second view – the *strong embodiment view* – argues that there are associative neuronal links between words and the experiences they refer to, suggesting therefore that the experiences themselves should be evoked by language. Indeed, Pulvermuller

and colleagues have proposed that action words are learned in the context of action performance and observation. Hebbian connections between words and action experiences are thus formed and give rise to distributed neural assembles linking words and sensory-motor information (Hauk, Johnsrude, & Pulvermuller, 2004; Pulvermuller, 2005). Specifically for motion-related activations in the posterior temporal lobe, Kiefer et al. (2012) have argued against the proposal that these activations are supra-modal, integrative or abstracted in nature. Instead, they argue for modality-specific representations selectively responding to motion or sound features of word meanings according their proximity to the corresponding auditory or visual sensory cortices (Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008; Kiefer et al., 2012). Similarly, other authors argue that motion sentences elicit visual simulations in V5 or recruit neural perceptual machinery, consistent with strong embodiment (Kaschak et al., 2005; Saygin, McCullough, Morana, & Emmory, 2010). On this view, the specificity and motion perceptual character of the representations elicited by language is greater than those schematic representations proposed by weak embodiment, and linguistic representations are more dependent on sensory experience, as in other strong embodiment accounts (Pulvermuller, Shtyrov, & Ilmoniemi, 2005).

The third view of action representations in temporal cortex is that these representations are independent of perception and therefore, unrelated to any motion content – a *modality-independent view* or *amodal view* (Bedny & Caramazza, 2011; Bedny et al., 2008). Bedny and Caramazza (2011) in particular, argue that all kinds of verbs, not just motion verbs, activate the posterior temporal lobe, whereas nouns with motion features do not. Moreover, congenitally blind patients show responses in this region similar to sighted participants, suggesting that the role of visual input in the formation of these representations is minimal, if any (Bedny, Caramazza, Pascual-Leone, & Saxe, 2012). Instead, they suggest that what may be critical to posterior temporal representations is the retrieval of event concepts or the grammatical information associated with verbs. On this view, therefore, the posterior temporal activations elicited by language are not modality-specific, have no specific motion content and no dependency exists between perceptual and linguistic representations of actions.

Previous research in motion language has partially supported all these views. It has been shown for example, that the activity elicited by motion verbs or sentences in temporal regions, compared to verbs and sentences referring to static events, barely overlaps with the identified V5 cluster (Revill, Aslin, Tanenhaus, & Bavelier, 2008) or is altogether located anterior to it in posterior temporal cortex, thus suggesting little activation of perceptual features in V5 (Bedny et al., 2008; Kable et al., 2002). Kable and colleagues also presented static images of moving objects (e.g., a train, a hand slicing bread, ducks walking) or words (*trailing, slicing, following*) and asked participants to judge the similarity of the concepts. They found that moving objects elicited activity both in V5 and posterior temporal cortex, whereas motion words elicited activity only in posterior temporal cortex and other language areas. More generally, numerous language studies using linguistic motion stimuli, but not necessarily comparing language and visual activity, have typically found motion or action sensitivity in posterior temporal cortex, but no hint of V5 activation (Chen, Widick, & Chatterjee, 2008; Gennari, MacDonald, Postle, & Seidenberg, 2007; Kellenbach, Brett, & Patterson, 2003; Noppeney, Josephs, Kiebel, Friston, & Price, 2005; Tyler et al., 2003; Wallentin, Lund, Ostergaard, Ostergaard, & Roepstorff, 2005; Wallentin et al., 2011).

Other studies however have shown V5 modulations as a result of linguistic content suggesting high sensory-specificity in the activations elicited by language. Saygin et al. (2010), for example, report higher activity in V5 when participants are exposed to audiovisual presentation of someone uttering motion and fictive motion sentences (*I drove from Modesto to Fresno* and *The highway runs from Modesto to Fresno* respectively), as compared to static sentences (*Modesto and Fresno are in California*). Participants had to indicate whether the sentences made sense. They found that motion sentences elicited more activity in V5 than fictive motion sentences and these in turn elicited more activity than static sentences, suggesting visual simulations in motion language processing (Saygin et al., 2010). In another study, Rueschemeyer and colleagues also report that language modulated activity in V5 (Rueschemeyer, Glenberg, Kaschak, Mueller, & Friederici, 2010). They manipulated the direction of motion: motion towards or away from the self or other people or objects (*the car drives toward you/Maria/the bridge, the car drives away from you*) (Kaschak et al., 2005). Results indicated that only movement towards the self elicited significant activation in the region defined as V5. The authors argue that the results reflect top-down modulations on V5 because motion towards the self requires the listener to track the objects' motion to evaluate potential subsequent behavior. This interpretation is consistent with many findings in the visual perception literature showing top-down attention modulations in V5 (Kourtzi & Kanwisher, 2000; O'Craven, Rosen, Kwong, Treisman, & Savoy, 1997; Senior et al., 2000). See (Gennari, 2012) for a detailed review of language and visual motion studies.

Yet other studies challenge the motion-specificity claimed by weak and strong embodied approaches above, particularly within posterior temporal regions. Indeed, posterior temporal regions respond to various kinds of linguistic stimuli with or without motion content: abstract mental verbs activate posterior temporal structures overlapping with those active for motion verbs (Bedny et al., 2008; Rodríguez-Ferreiro, Gennari, Davies, & Cuetos, 2011) and noun–verb ambiguous words like *hammer* more strongly activate these regions when used as verbs, compared to nouns, suggesting a role for verb semantics and inflexional morphology (Gennari et al., 2007; Tyler, Bright, Fletcher, & Stamatakis, 2004; Tyler, Randall, & Stamatakis, 2008). Moreover, these regions also respond to nouns in situations in which processing is difficult, e.g., when deciding which of two nouns is related to a third. This has led some authors to argue for general semantic control or selection mechanisms in this region (Whitney, Jefferies, & Kircher, 2011; Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2011, 2012). Therefore, as argued by Bedny and Carammazza, it remains unclear the extent to which motion content *per se* is indeed critical for posterior temporal activations.

Here, we re-examine these issues and ask two specific questions: (a) whether descriptive language re-activates low-level motion representations such as those encoded in V5, in addition to those in posterior temporal cortex, and (b) whether motion content characterizes activations in posterior temporal cortex, thus suggesting modality-specific representations. To this end, we used descriptions of visual scenes as stimuli and compared sentences describing motion events involving humans and objects to corresponding sentences describing static events. The visual scenes encouraged visual imagery during language processing but critically, they were not present during language comprehension. For this, we adapted the blank-screen paradigm widely used in psycholinguistic eye-tracking studies (Altmann, 2004; Altmann & Kamide, 2009; Knoeferle & Crocker, 2007; Richardson & Spivey, 2000). In these studies, participants are first shown an image of a scene. The image is then removed and replaced by a blank screen. Participants then hear a sentence that refers back to the previously viewed scene. During this time, participants interpret the language relative to the mental representation of the scene and thus move their eyes to the location in which objects had been, despite the scene being no longer present.

In our language task, participants saw a scene such as that of Fig. 1 for 2 s. After a variable period (between 2 and 4 s), they heard
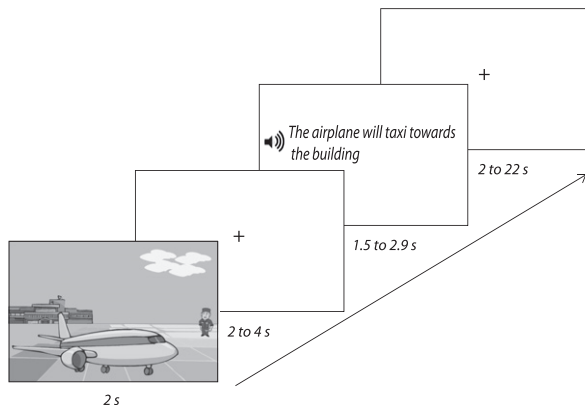
**Fig. 1.** Trial structure from the language task.

sentences varying in motion and animacy (human or object motion vs. human or object static sentences). Participants were instructed to fixate on a cross during this time and to press a button only when the sentence did not refer to the picture, thus encouraging the processing of both the sentence and the scene. Example sentences are below:

1a. The mechanic will walk towards the airplane (human motion).
1b. The mechanic is looking at the airplane (human static).
1c. The airplane will taxi towards the building (object motion).
1d. The airplane is stationed near the building (object static).

This task was preceded by a localizer scan, which served to identify the brain regions targeted in our analyses such as area V5 and posterior temporal cortex.

By adopting this procedure, we aim to improve previous methodologies. First, we reasoned that if there are perceptual simulations or highly modality-specific representations elicited by language, the strongest test for such representations is the use of linguistic stimulation separate from visual stimulation, unlike Saygin et al. (2010)'s procedure, because the words themselves should be able to activate visual experiences without need of concurrent visual input. Second, the procedure encourages visual imagery by asking participants to refer the sentence to a mental image of the scene. Unlike the metalinguistic judgement tasks used in previous studies, this mimics natural situations in which we talk about events without necessarily being confronted with them, at the same time that it improves the chances of finding an effect. Finally, the procedure encourages relatively constant attention to perceptual memories across experimental conditions and does not include references to the self, as in Rueschemeyer et al., 2010's study, eliminating potential top-down influences based on emotional content, such as fear or avoidance of approaching objects.

We tested the main predictions of the theories outlined above. First, if language elicits schematic representation of events that are nevertheless motion-specific (i.e., modality-specific), as suggested by weak embodiment but not modality-independent views, then, both human or object motion sentences should elicit stronger activity than their corresponding static counterparts in posterior temporal regions also responding to human or object visual motion respectively, rather than in area V5 proper, whereas little difference should be observed between the static conditions. Second, if language processing recruits low-level modality-specific representations, as argued by strong embodiment and demonstrated in Saygin et al. (2010), then it might also modulate activity in motion area V5, which is known to respond to both low-level human and object motion (Beauchamp et al., 2002). To test these predic-

tions, we contrasted motion vs. static sentences within individually-defined V5 regions. This is particularly important because the location of this area varies greatly across people and may give rise to unwanted biases when conducting group level analyses in standard space (Dumoulin et al., 2000; Swallow, Braver, Snyder, Speer, & Zacks, 2003). Within posterior temporal structures, we examined activity elicited by our sentence conditions in regions of interest responding to either human or object motion, as obtained from the group-level localizer scan.

## 2. Experimental methods

### 2.1. Participants

Ten right-handed native English speakers (three male and seven female, mean age = 23.8 years) with no history of psychiatric illness or neurological damage were scanned. Note that this number of subjects is appropriate, particularly if positive results are found – see recent discussion by Friston (2012).

### 2.2. Materials

#### 2.2.1. Visual stimuli

The localizer task included 39 motion video clips lasting 4 s paired with corresponding static images. Three stimulus types were contrasted with corresponding static stimuli to localize V5: radially expanding and contracting dot and grating patterns, human motion videos and object motion videos. Moving dots and gratings have been extensively used in the literature to identify area V5 (Beauchamp et al., 2002; Dumoulin et al., 2000; Tootell et al., 1995). The object stimuli consisted of self-propelled vehicles such as planes, trains, trucks and cars that were also used in the sentence stimuli. Example stimuli are shown in Fig. 2.

#### 2.2.2. Language stimuli

Fifty pictures of scenes (Fig. 1) were paired with four different sentence conditions as exemplified in (1) above. All figures and objects in the scenes were in a resting position so that the same scene was used for both static and moving conditions. This feature of the design guaranteed that differences across the sentence conditions were not due to differences in the preceding scene. The sentences varied in terms of motion (motion vs. static) and moving entity (human or object) and took the form subject-verb-prepositional phrase in third person singular (except for a few cases in which they had subject-verb-object structure). As exemplified above, motion sentences were in future tense whereas static sentences were in present tense. This guarantees that the motion content cannot be easily inferred from the picture and encourages the simulation of a soon to occur event in the context of the picture.

There were 200 sentences in total (100 motion sentences and 100 static sentences, divided equally between human and object motion or static content). The full stimulus set is provided in the Appendix. Sentence length, and verb frequency were matched across conditions (Cobuild Corpus: 20 million words). Mean and standard deviations for these variables are shown in Table 1 along with two examples. T-tests across conditions showed no significant differences in word frequencies or length. Note that the nouns are the same across the motion and static conditions and therefore they were matched for lexical differences. The stimulus sentences were recorded by a male native English speaker who spoke naturally and clearly. After recording, the audio files were normalized to the same intensity.

*2.2.2.1. Motion imagery pre-test.* To check that the motion and static conditions differed in terms of motion imagery as intended, we
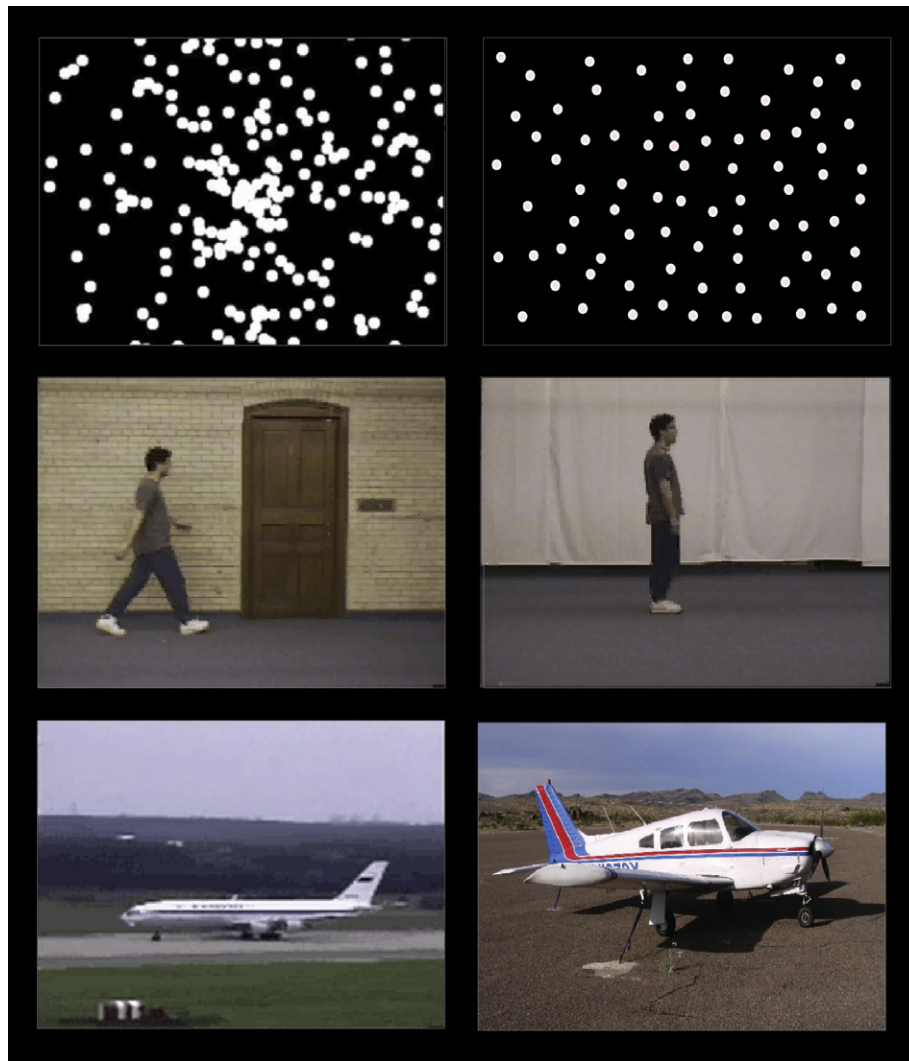
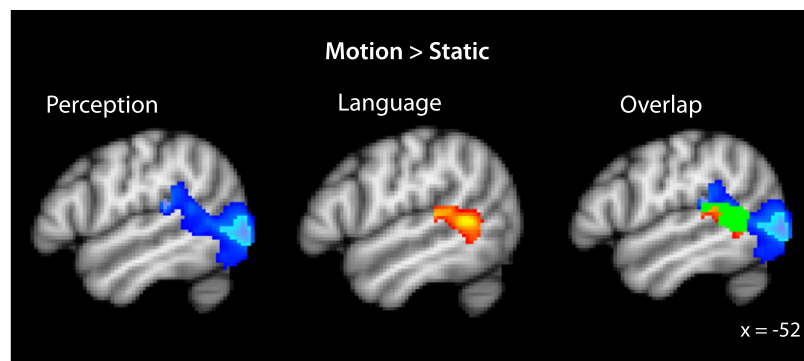**Fig. 2.** Example videos from the perceptual localizer task.



**Fig. 3.** Results from perceptual localizer and language tasks (computed in standard space using high-level statistics). Activity in blue indicates the results of the perception localizer, with light blue indicating the strongest cluster responding to motion (area V5). Activity in red shows the results of the language task. Green indicates the region of overlap between perception and language.

conducted a web-based questionnaire. 42 participants were asked to rate each sentence in a scale of 1–7 indicating the degree to which the sentence brings to mind images or experiences of motion within the context of the scene they were paired with. The questionnaire was organized in 4 lists, each containing only one sentence version associated with a picture but all conditions across

items. Motion sentences received much higher ratings than static sentences, but there was a small difference between object motion and human motion ratings with no difference between the static sentences (see Table 1). A repeated measures ANOVA with mean rating *per se*ntence as dependent variable and motion and animacy as main factors revealed that there was a main effect of motion

**Table 1**
Language stimulus characteristics.

| | Examples | Mean log frequency | | | Mean length (ms) | Motion Imagery |
|---|---|---|---|---|---|---|
| | | Noun 1 | Verb | Noun 2 | | |
| Human motion | 1. *The milkman will approach the house* <br> 2. *The farmer will open the gate* | 4.15 (.62) | 4.52 (.77) | 4.22 (.71) | 2.11 (.25) | 4.83 (.65) |
| Object motion | 1. *The milk float will drive up the road* <br> 2. *The tractor will plough the field* | 4.11 (.62) | 4.59 (.60) | 4.26 (.59) | 2.07 (.25) | 5.31 (.59) |
| Human static | 1. *The milkman is resting near the road* <br> 2. *The farmer is looking over the gate* | 4.15 (.62) | 4.66 (.67) | 4.22 (.71) | 2.14 (.27) | 2.20 (.58) |
| Object static | 1. *The milk float is parked near the road* <br> 2. *The tractor is stuck in the field* | 4.11 (.62) | 4.61 (.52) | 4.26 (.59) | 2.09 (.26) | 2.23 (.73) |

*Note*: Mean values are provided with standard deviations in parenthesis.

$(F(1, 49) = 1260, p < .0001)$, a main effect of animacy, $(F(1, 49) = 6.44, p = .01)$ and an interaction $(F(1, 49) = 7.55, p = .008)$. Object motion sentences thus elicited more motion imagery, perhaps because objects such as trains and plains in these sentences move at higher speed than humans.

### 2.3. Experimental design

#### 2.3.1. Localizer design

The localizer task had a block design. Each block contained three videos of moving stimuli or three static images presented for 4 s each. "OFF" periods consisted of a white fixation cross on a black background. There were 26 blocks of moving videos or static pictures (baselines) presented in a randomized order with "ON" and "OFF" periods of 12 s (total duration = 10 min 25 s). In this task, participants were instructed to watch the videos and try to remember the events depicted on them for a subsequent test after scanning. They were also instructed to keep fixating on the middle cross at all times, both while the videos played or in rest periods.

#### 2.3.2. Language task

This task used an event-related design (Fig. 1) and took place after the localizer task. In each trial, a picture of a scene was presented for 2 s. The scene was then removed and replaced with a fixation cross that remained on the screen until the end of the trial. After a short interval (randomly varied from 2 to 4 s) an auditory sentence was presented at a volume of 50 dB (mean duration = 2.1 s, SD = .26). The inter-trial interval randomly varied from 2 to 22 s. There were a total of 242 trials, each containing a picture and sentence. The language task lasted a total of 42 min and was divided in two continuous sessions of 21 min each. On the majority of trials (77%), the sentence referred back to the previous scene. In the remaining *catch* trials, this was not the case. Participants were instructed to fixate on the cross while it was on the screen and to indicate by a button-press when the content of the sentence *did not* refer to that of the picture (catch trails).

The sentences were presented in a pseudo-randomised order whereby each condition followed every other condition with an equal probability. The stimulus order and inter-trial interval durations were obtain using a scheduler that computes the most efficient durations and condition order to estimate the hemodynamic response for each trial and event type (http://surfer.nmr.mgh.harvard.edu/optseq). To minimize order effects, the experiment was divided into two parts, and the order in which each part was presented was counterbalanced across subjects. Moreover, the four different sentences associated with a picture (i.e., human and object motion and human and object static) were presented in variable orders across the experiment so that in total there were 22 different sentence orders associated with a picture for each subject, and these orders were different across subjects due to the counterbalancing of the first and second part of the experiment. These features of the design therefore guarantee that there was little repetition of the order of sentence types associated with a given picture and make it very unlikely that a particular order influenced the average activity across sentences in one condition.

### 2.4. Scanning parameters

Imaging was carried out using a 3 T GE Signa Excite MRI scanner. High-resolution whole brain structural images were obtained for each participant. Functional images were obtained using a gradient-echo EPI sequence (TR 2000 ms, TE 50 ms, flip angle 90°, matrix $64 \times 64$, field of view 21 cm) with 30 axial slices of thickness $3 \times 3.28 \times 3.28$ mm. Functional images excluded the dorsal part of the motor cortex. A high-resolution T1 Flair image was also obtained to aid in co-registration.

### 2.5. Data analysis

#### 2.5.1. Standard space analyses

Data analyses were carried out using FSL tools (the software library of the Oxford Centre for Functional MRI of the Brain (FMRIB); www. fmrib.ox.ac.uk/fsl). First- and higher-level analyses were carried out using FEAT (FMRI Expert Analysis Tool). The two language sessions were combined by concatenating the demeaned raw sessions' data (an alternative procedure, combining parameter estimates for each individual session, yielded virtually identical results). Pre-processing of the data (both for language and localizer tasks) included motion correction (Jenkinson, Bannister, Brady, & Smith, 2002), slice-timing correction, brain extraction, spatial smoothing using a Gaussian kernel of full-width-half-maximum (FHWHM) 8 mm, and high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, sigma = 25.0 s). Motion correction indicated that participants did not move more that 0.10 mm on average.

Time series analyses were conducted using FILM (FMRIB's Improved Linear Model) with local autocorrelation correction (Wool-rich, Ripley, Brady, & Smith, 2001). The *localizer data* were modeled with the standard box function of block designs convolved with a hemodynamic response function (gamma function). For these data, we used the contrast motion videos > static pictures to identify area V5, which included all stimulus types – dots and gradients, human and object stimuli – because there was little difference between these stimulus types in this region, consistent with previous reports. The *language data* were modeled with event-based explanatory variables corresponding to the condition and trial structure, which were convolved with a hemodynamic response function (gamma function). Picture and sentence events within a trial were modeled separately. One variable modeled all pictures together, as these pictures were always the same across the sentence conditions. The sentence event was modeled according to the sentence

conditions: human motion, object motion, human static and object static and catch trials. In the model, each HRF was aligned with the beginning of each picture or sentence event and lasted for the duration of the picture or sentence. We focused our V5 analyses on the main overall motion contrast motion > static (assigning 1 to human and object stimuli and −1 to the static versions) to gain statistical power, given that all motion sentences, like all moving stimuli, should elicit activity in V5. For the analyses in the posterior temporal lobe, we also examined the activation level of each condition compared to rest in region of interest analyses.

For the group results using standard higher-level statistics, we report analyses conducted within a left-hemisphere from which auditory cortex (defined by Juelich histological atlas built into FSL) was removed. This was done because the sentences contained different words, which were likely to produce differential effects in auditory cortex of no interest in the present study.

### 2.5.2. Region of interest analyses in standard space

We also report two region of interest (ROI) analyses, one in V5 and another in the posterior temporal lobe, using areas defined by the localizer contrasts in the standard brain. For ROI analysis in V5, we identified the strongest clusters responding to all motion videos in the localizer task, and ran a high-level one-sample t-test across the group to evaluate whether language modulated activity in this region. For the ROI analysis in the posterior temporal lobe, we identified the posterior temporal region responding to both visual motion and motion language and then we extracted the signal percent change within this cluster for each of the language conditions (human motion, human static, object motion, object static). These values were further submitted to a standard analysis of variance to examine modulations across conditions. This latter analysis examined whether motion-specific content elicited strong language responses, as suggested by weak embodiment.

### 2.5.3. Region of interest analysis in each individual brain

Although a high-level analysis may not yield any significant difference for a given contrast, true differences may be overlooked due to group averaging or transformation to standard space – a process that introduces significant noise (Juch, Zimine, Seghier, Lazeyras, & Fasel, 2005) – or simply due to individual differences in the areas sensitive to motion (Swallow et al., 2003). The location of V5 has indeed been shown to vary greatly across individuals and to influence group inferences using common high-level analyses procedures in standard space (Dumoulin et al., 2000; Swallow et al., 2003). We thus aimed to obtain more precise activations by looking at regions of interest defined in each individual brain. Two left-hemisphere ROIs were defined from the localizer data at the individual level. A ROI sensitive to motion perception in area V5 was identified in each individual by inspecting the motion >- static localizer contrast and following the anatomical landmarks delineated in previous work (Dumoulin et al., 2000). The posterior temporal ROI was also defined from the motion > static localizer contrast and it was the cluster located anterior and lateral to V5. For each individual, the ROI was manually drawn around the cluster of voxels that showed the strongest activity in the V5 area or the posterior temporal area. So that no effects were missed, we aimed to include relatively large clusters containing between 120 and 240 voxels. This variability in the cluster sizes across subjects is due to individual differences in the cluster sizes emerging from the localizer contrast. For subjects that had relatively small clusters of activation, we lowered the contrast image threshold to z = 2 so that sufficiently large clusters could be drawn. See examples of ROI masks in Fig. 7. We then computed the motion vs. static contrast from the language data within these ROIs. Z (Gaussianised T/F) statistic images were thresholded using GRF-theory-based maximum height thresholding with a (corrected) significance

threshold of P = 0.05 (Worsley, 2001). Finally, we made group-level inferences from these individual ROI analyses by conducting two types of statistics. First, we used chi square statistics to evaluate whether the probability of finding significant effects in each ROI is greater than that expected by chance. Second, we entered each participant's mean percent change calculated from the motion vs. static contrast parameter estimates into one-sample t-tests (with subject as random factor) to evaluate whether the group shows significant activity increases within each ROI.

## 3. Results

### 3.1. Behavioral performance in the scanner

Behavioral performance in the language task was high for all participants (mean% correct = 94.6%, SD = 2.8). The mean percentage of false alarms was low (3.8%), and thus these trials are unlikely to influence the overall results. This indicates that participants remained alert throughout the experiment and correctly identified those trials in which the sentence did not refer to the previously seen visual scene.

### 3.2. Left occipito-temporal activations in language and perception (standard space analyses)

Fig. 3 shows the responses to motion (motion > static) for the localizer and language tasks from the group analysis within the left hemisphere. The clusters in dark blue represent the significant voxels in the perception localizer (cluster corrected, z = 2.3, p = .05) and the areas in red show the significant voxels in language (cluster corrected, z = 2.3, p = .05). The light blue cluster shows the strongest voxels responding to this contrast (thresholded at z = 3.2), which is the cluster corresponding to area V5 proper (Dumoulin et al., 2000). Table 2 shows activation peaks. Motion perception elicited responses in superior and middle temporal gyrus, inferior parietal lobe and lateral occipital cortex, whereas motion sentences compared to static sentences elicited responses in a portion of the posterior middle and superior temporal gyri. Importantly, when the strongest responses to motion in area V5 are selected for an ROI analyses in the standard space (the area represented in light blue in Fig. 3), no significant language activations emerged at the group level (voxel based corrected at p = .05).

These results are consistent with weak embodiment: motion sentences and the perception of motion elicited shared responses within left posterior middle and superior temporal gyri, as shown by the green area of common activation in Fig. 3. In contrast, they are not consistent with the strong embodiment views, as no overlap was found between language and motion perception area V5.

### 3.3. Region of interest analyses within temporal cortex (standard space)

To examine whether the language activity in posterior temporal structures is specifically due to motion content, as argued by weak embodiment, we extracted the percent signal change for each lan-

**Table 2**
Peak centers in MNI space for contrasts performed in the language and perception tasks.

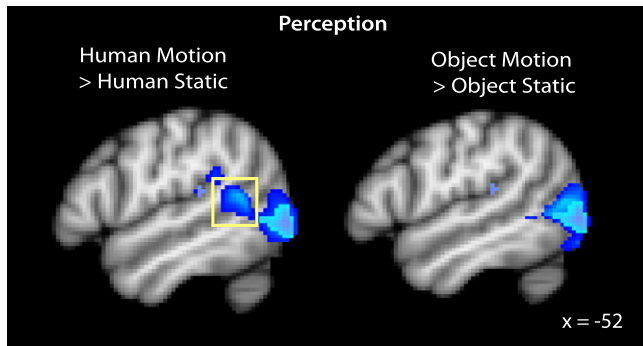| Contrasts | Brain area | Perception | Language |
|---|---|---|---|
| Motion > static | Middle occipital gyrus | −48, −76, −4 | |
| | Post-central gyrus | −52, −34, 18 | |
| | Superior temporal gyrus | −54, −48, 14 | |
| | Middle temporal gyrus | | −54, −56, 6 |

**Fig. 4.** Localizer results for human and object motion. The area highlighted within the yellow square overlaps with the overall motion results in Fig. 3 shown in green.
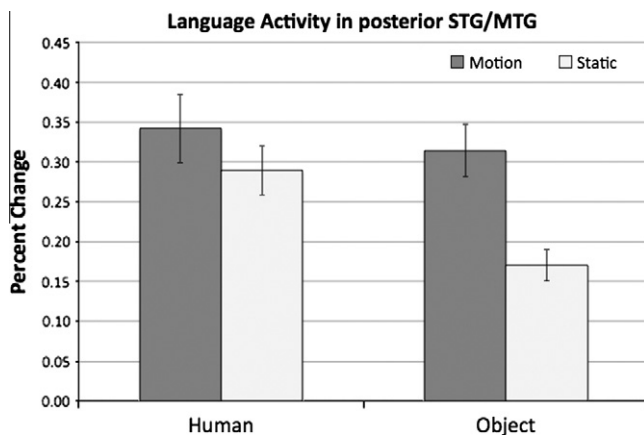


**Fig. 5.** Average percent signal change for each condition relative to rest within the posterior temporal ROI.

guage condition within the area of language and perception overlap shown in green in Fig. 3. For completeness, we also checked the language activity within the localizer region that responded to human motion specifically, rather than general motion, but the pattern of results was the same because these two regions of interest largely identified the same cluster. As shown in Fig. 4, the contrast human motion > human static revealed an area of the posterior superior temporal gyrus (highlighted in yellow in Fig. 4), which largely overlapped with the temporal region responding to overall motion in Fig. 3. This finding is in line with numerous perceptual studies that have shown sensitivity to biological motion in this region (Beauchamp et al., 2002; Blakemore et al., 2003; Morito, Tanabe, Kochiyama, & Sadato, 2009; Puce & Perrett, 2003; Schultz, Friston, O'Doherty, Wolpert, & Frith, 2005a; Servos, Osu, Santi, & Kawato, 2002; Thompson, Clarke, Stewart, & Puce, 2005). Interestingly, the localizer contrast object motion > object static did not yield any specific cluster beyond that surrounding area V5 (see Fig. 4). Previous studies have argued that perceptual object representations are encoded in the middle temporal lobe (Beauchamp, Lee, Haxby, & Martin, 2003; Beauchamp et al., 2002), but these studies have only investigated tool motion, rather than vehicle motion as in the present study.

The results of these ROI analyses (both within the language-perception overlap (Fig. 3) and the human motion localizer region (Fig. 4)) revealed effects of the specific language conditions: both motion and animacy elicited higher levels of activation compared to object static sentences. The percent change values for each condition averaged across participants are shown in Fig. 5 (where error bars were computed as suggested by Loftus and Masson (Loftus & Masson, 1994)). A repeated-measures ANOVA with percent change as dependent variable and sentence motion and animacy as factors showed a main effect of animacy ($F(1, 9) = 18.51$, $p = .002$), a main effect of motion ($F(1,9) = 28.96$, $p < .0001$) and an interaction ($F(1, 9) = 9.84$, $p = .01$). There was a significant difference between human motion and human static sentences ($p = .03$) and between object motion and object static sentences ($p < .0001$), but the effect sizes were different (motion had a larger effect on
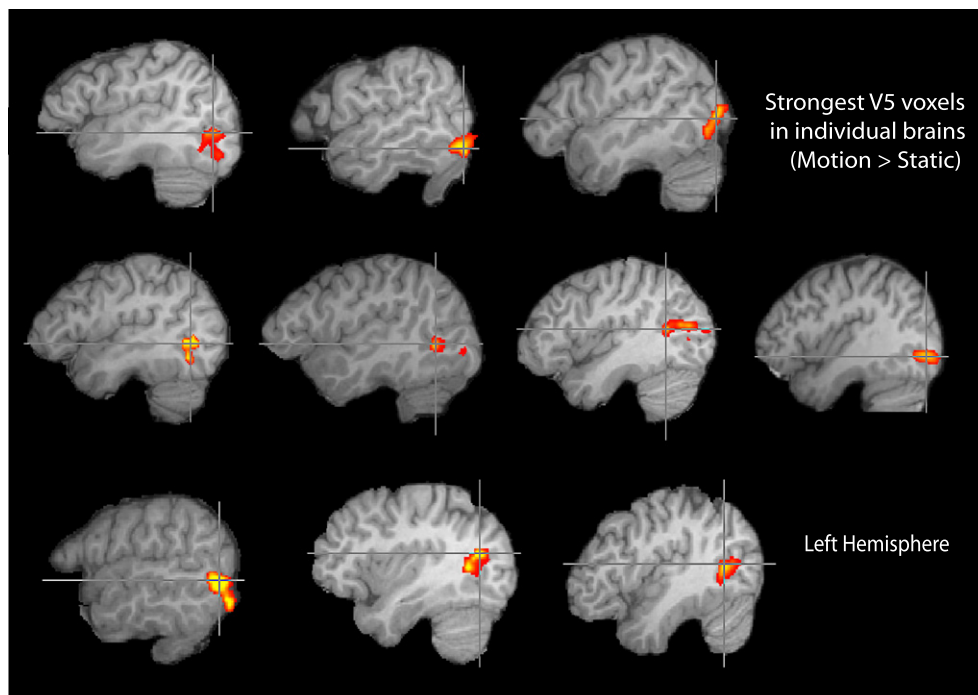


**Fig. 6.** Strongest voxels responding to visual motion stimuli in each individual brain in the left hemisphere. Each brain is displayed in individual structural space.

objects), hence the interaction. There was no significant difference between human and object motion. These results clearly suggest that the language activity does not specifically respond to motion, but also reaches high levels of activation for human static sentences, though not as high as that for motion sentences. Thus, it appears that posterior temporal activations in language not only respond to motion but also to other aspects of events, as suggested by the modality-independent view. In this case, having animate participants and verbs that imply intentionality (e.g., *looking, waiting, thinking*) increases activations in posterior temporal structures. We will discuss the implications of this finding for the theories outlined in the introduction in the discussion.

### 3.4. Region of interest analyses in individual brains (motion vs. static sentences)

The results reported above for area V5 were conducted in standard space so that it is possible that brain normalization and averaging techniques masked true effects within V5 in individual participants. The more accurate ROI analyses conducted in each individual brain revealed that the location of the strongest voxels responding to motion in the localizer data indeed varies greatly across participants, as exemplified in Fig. 6, making the present

individual analyses worthwhile. Some participants have more lateral rather than medial clusters, whereas some show two clusters, rather than one. Table 3 also provides the MNI coordinate of each participant's V5, where the maximum voxel of activation in individual space was transformed into standard space using individual-specific transformations.

The language activity (motion > static) extracted for each individual V5 ROI (defined as specified in Section 2.5.3) was found in only four of the 10 participants tested and only at the very anterior extent of the region identified as V5 (voxel corrected, $p < .05$). A chi square test indicated that the probability of finding an effect across the group within V5 was not significantly larger than that expected by chance ($\chi^2 = 2$, $p = $ n.s.). Fig. 7 shows the results for two representative individual participants. Participant 5 is one of the participants that showed a small effect in the area defined as V5, as well as an effect in the posterior temporal cluster. Participant 9 exemplified the majority of cases in which no activity was found within V5. This variability in participants' activations may be due to individual differences cognitive styles, e.g., some participants engage in more visual imagery than others, as suggested by similar findings in the domain of color processing and color naming (Hsu, Kraemer, Oliver, Schlichting, & Thompson-Schill, 2011). Note that even applying the most lenient uncorrected statistic to the language
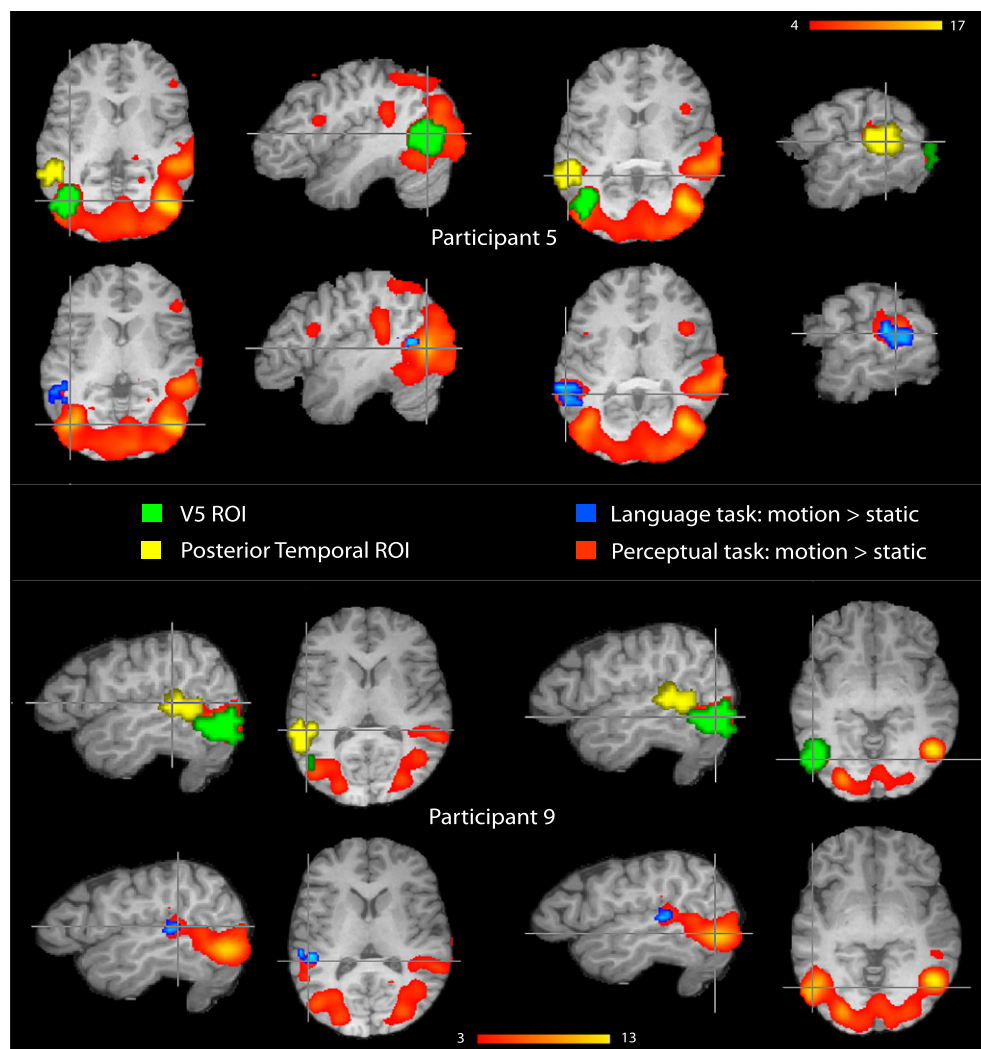


**Fig. 7.** Examples of region of interests and data from the analyses in each individual participant. The top panel displays results from an experimental participant showing activity within the cluster defined as V5, whereas the bottom panel displays the most common pattern of activation found across participants. The top row images in each panel show the masks that were used as regions of interest for each participant, and the bottom row shows the overlay of the language and the perceptual activity.

**Table 3**
MNI coordinates for each participant's maximum voxel within V5 (left hemisphere).

| Participant | MNI coordinate |
| --- | --- |
| 1 | −40, −70, −4 |
| 2 | −44, −72, −8 |
| 3 | −52, −72, −8 |
| 4 | −40, −72, 10 |
| 5 | −50, −76, −16 |
| 6 | −46, −72, −2 |
| 7 | −46, −72, 4 |
| 8 | −50, −72, −4 |
| 9 | −38, −68, 4 |
| 10 | −44, −84, −12 |

data (uncorrected, $p = .05$) results in little activity observed within area V5 in nearly all participants, thus suggesting that the variability we found is not due to stringent correction methods.

To further explore the activation levels within V5 irrespective of whether this activity was significant for a given individual, we obtained the percent signal change for each participant's motion vs. static contrast and conducted a one sample t-test across the group. This analysis showed that the mean activity within V5 was (marginally) significantly smaller than 0 in a two-tailed test ($t(9) = -2.18$, $p = .06$), suggesting de-activation in this region as a whole, rather than positive activation. See Fig. 8. Thus, despite the fact that some participants showed some significant voxels within V5, the overall level of activation within this region is low.

To compare these results in V5 with posterior temporal results, similar analyses were conducted within the posterior temporal ROI. These yielded similar results to those reported in the high-level standard analysis. All subjects, except for one, showed significant differences between motion sentences and static sentences in the posterior temporal ROI. Across the group, therefore, a chi square test indicated that the probability of finding an effect within posterior temporal cortex was significantly larger than that expected by chance ($\chi^2 = 32$, $p < .0001$). Moreover, a one sample t-test using the percent signal change of each participant's motion vs. static contrast as dependent variable indicated that the group showed effect sizes larger than 0 ($t(9) = 6.47$, $p < .0001$). See Fig. 8. These results thus confirm the presence of a motion effect
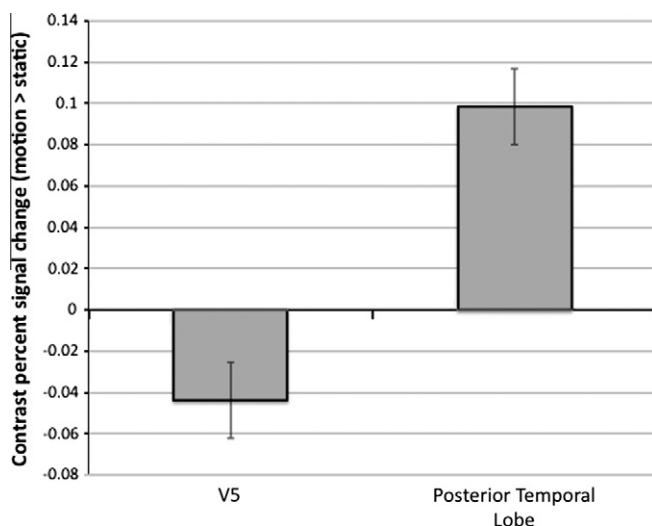


**Fig. 8.** Average percent signal change computed from the motion > static contrast parameter estimates across individuals. Each participant's data point was obtained from the regions of interest (V5 and posterior temporal clusters) defined in each individual brain.

in temporal cortex, consistent with that reported above, and furthermore indicate that the positive effect size in posterior temporal cortex is much larger than that in V5.

Taken together, these results provide little evidence that descriptive language modulates activity within motion perception area V5, despite the fact that some individuals show an effect. First, the study had sufficient power to observe large effects in posterior temporal cortex whereas no effect is observed in V5. Second, even if a small effect is found in this region (for example, by testing many more subjects), the overlap of language with V5 activity would most likely be partial and located in the anterior portion of this region, as shown in the present data, leaving the strongest voxels responding to motion perception outside the reach of language. Indeed, the positive individual effects found here are due to our attempt to extend the size of this region as much as possible, since alternative analyses that we have conducted with smaller cluster sizes (the 80–100 strongest V5 voxels) did not yield any significant language result. Therefore, it is clear that activity within V5, if any, would follow a posterior–anterior gradient, with more posterior regions only responding to perception and more anterior multimodal regions responding to both perception and language.

## 4. Discussion

One goal of the present study was to examine the extent to which language recruits low-level perceptual representations of motion, using a task known to elicit language interpretation with reference to a visual scene (Altmann, 2004). We found that in occipito-temporal cortex, language did not typically recruit low-level motion perception areas such as V5. This was shown both with analyses conducted on ROIs defined in the standard space and with ROIs defined separately for each individual participant. Although some participants showed some activity within a loosely defined V5 cluster, this activity was only found in the more anterior portion of the cluster and was not sufficient to generate significant group results. However, in most participants language did recruit posterior temporal regions that were also active when observing motion events. These results therefore suggest that sensory-based representations in language do not have the specificity found in visual perception, even when reference to a visual scene was encouraged. We discuss the implications of these findings for theories of semantic representations below.

Another goal of this study was to determine the nature of the semantic content recruited by language in posterior temporal regions, and specifically, whether these respond to motion only. We did find that the posterior temporal cluster active in language comprehension show a main effect of motion in all analyses. However, closer inspection of the level of activation elicited by each of the language conditions revealed that whereas motion sentences indeed elicited the strongest activity, human static sentence also did so, albeit to a lesser degree, hence the interaction shown in Fig. 5. This finding indicates that posterior temporal activity not only responds to motion but also to other aspects of event structure, such as animacy and/or the intentionality of the action described. We discuss the implications of these findings below.

### 4.1. Sensory-specificity of semantic representations in language

Our motion results in V5 and posterior temporal cortex provide support for weak embodiment theories of meaning. Although motion sentences recruited areas of the brain that were involved in motion perception, this was the case only within posterior temporal cortex (Figs. 3 and 7). These findings are in line with many language studies reporting similar motion-related regions and no activity in V5 proper (e.g. Kable et al., 2002, 2005; Bedny et al.,

2008; Revill et al., 2008; Wallentin et al., 2005). In contrast, activity in V5 was only observed in a few participants in the more anterior portion of the V5 cluster, and it did not reach significance at the group level. This is *in*consistent with strong embodiment views (e.g., visual simulations), although some individuals *can* activate this region, possibly because they engage in more mental imagery (Hsu et al., 2011). This pattern of results is therefore consistent with the proposal that there is a concrete-abstract motion gradient through the lateral occipito-temporal cortex (e.g. Kable et al., 2002; Martin & Chao, 2001) and suggests that language effects on V5 do not appear to be present as part of the comprehension process when merely listening to a description of a motion event.

Nevertheless, the fact that previous studies have found effects within the region defined as V5 indicates that such activations may occur in some circumstances. First, as previously shown, V5 is sensitive to attention demands and responds more strongly to moving stimuli when participants are told to attend to it (O'Craven et al., 1997). It is therefore possible that certain tasks and stimuli focus participants' attention on motion features more than others. This may have been the case in some of the previous studies, in which for example, strong emotional content, as when objects move towards the self (Rueschemeyer et al., 2010), may elicit top-down effects on motion perception. Similarly, Saygin et al. (2010) used concurrent visual stimulation that required audio-visual integration of the talking face with the auditory sentences. This process may elicit additional feedback between posterior temporal cortex, which is known to perform audiovisual integration, and area V5 (Calvert, Campbell, & Brammer, 2000; Calvert et al., 1999). Moreover, methodology varies greatly across studies, so that an area V5 loosely defined may include portions of the posterior temporal lobe and therefore yield significant language activity, as demonstrated here and other previous studies (Revill et al., 2008). See Gennari (2012) for a detailed methodological discussion.

In our study, we tried to circumvent some of these methodological features, by omitting audiovisual integration and avoiding strongly focused attention on potentially emotive stimuli, while at the same time encouraging motion imagery. We also tried to target language uses that mimic situations in which people would report an event recently observed or about to happen. This seemed to us a stronger test of strong embodiment theories compared to previous studies, as perceptually-based representations should be elicited in the absence of visual motion stimulation and emotional content. Of course, if linguistic descriptions are concurrent with visual stimuli, it is certain that activity in V5 would be found (Tranel, Martin, Damasio, Grabowski, & Hichwa, 2005), as audiovisual integration must take place (Calvert et al., 1999; Calvert et al., 2000). As previously suggested, sensory-motor representations are flexibly recruited as a function of task demands and contexts (Binder & Desai, 2011). Our results therefore suggest that in circumstances in which descriptions of events are understood with reference to visual memories, V5 is not very strongly engaged, although there appears to be individual difference in how much mental imagery is elicited.

### 4.2. Semantic representations in posterior temporal cortex

Our results and others previously reported provide clues as to the exact nature of the representations entertained in temporal cortex, which responds to both linguistic and visual stimuli. Weak embodied approaches investigating motion language have argued for graded representations going from specific visually-based representations in V5 to more schematic motion representation in posterior temporal cortex, but weak embodied claims have not been thoroughly investigated with other types of non-motion stimuli. In the present study, we find that not only motion sen-

tences but also human static sentences strongly activate posterior temporal cortex. This suggests that motion per se is not a necessary component of these activations, as suggested by Bedny and colleagues, but rather, that additional properties of the event such as intentionality and animacy are computed here. This finding is consistent with many other language studies suggesting that posterior temporal regions, and particularly the superior temporal gyrus, encode event knowledge such as who might do what to whom and with what purpose (Bedny et al., 2008; Bornkessel, Zysset, Friederici, von Cramon, & Schlesewsky, 2005; Grewe et al., 2007; Peelen, Romagno, & Caramazza, 2012; Rodríguez-Ferreiro et al., 2011). Critically, this finding is also consistent with perception studies showing that these regions do not only respond to biological motion, as found here, but also to intentionality and animacy (Blakemore et al., 2003; Morito et al., 2009; Schultz, Firston, O'Doherty, Wolpert, & Frith, 2005b), an aspect that was not captured by our localizer scan. For example, Schultz, Friston, O'Doherty, Wolpert, and Frith (2005) parametrically varied the degree to which two circles on the screen appeared to chase each other, increasing the perception of animacy and goal-directed motion. They found that as the perception of animacy increased, so did activity in superior temporal regions. Thus, the level of schematization that appears to take place in posterior temporal representations goes beyond motion content and includes more schematic event representations that are nevertheless common to vision and language. Such multi-modal regions are often construed as convergence regions (Binder & Desai, 2011; Damasio, 1989) that bind information received from other regions into coherent representations. These observations therefore are consistent with weak embodied approaches in that a dependency of linguistic representations on low-level sensory-motor representations is maintained, even if they are schematized over the course or learning and/or bound together in multi-modal representations (Barsalou, 2003; Barsalou et al., 2003).

The evidence so far however is not consistent with strong embodiment views according to which sensory-motor experiences, perceptual simulations or modality-specific motion features are recruited by language (Hauk et al., 2004; Kaschak et al., 2005; Kiefer et al., 2012; Pulvermuller, 2005; Saygin et al., 2010). This evidence instead suggests that such views need to be further qualified, as language comprehension does not seem to engage modality-specific content but supra-modal conceptual representations that are shared with other modalities. Nevertheless, further studies are required to investigate these issues because there might be subdivisions within posterior temporal cortex, and both modality-specific and more abstract representations may turn out to display selective and segregated activity, less consistent with the graded continuum proposed by weak embodiment.

The evidence so far also suggests that a modality-independent view of posterior temporal activations may not be entirely correct. Although conceptually difficult tasks with nouns do activate temporal cortex and thus challenge the event-based semantic content proposed here (Whitney, Jefferies et al., 2011; Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2011; Whitney et al., 2012), it remains unclear whether these tasks engage action-based knowledge or affordances associated with the stimulus objects, as tool-related stimuli do, which would be consistent with event-based representations in posterior temporal cortex. Similarly, the evidence presented by Bedny et al. (2012) concerning blind individuals, as acknowledged by the authors, is consistent with weak embodiment views in that information from modalities other than vision may converge in posterior temporal structures, especially because re-organization of neural pathways may take place in such individuals. Given that in typically developing individuals, visual and linguistic (auditory or visual) information converge in posterior temporal cortex, it is likely that event-based information ab-

stracted from experience plays a role in language-induced representations. Thus, there may still be some dependency between modality-specific input and abstracted event knowledge recruited in language processing.

Finally, note that the claim of event-based representations in posterior temporal cortex is consistent with grammatical information also playing a role. Sentence comprehension studies typically involve posterior temporal regions (Bornkessel et al., 2005; Tyler & Marslen-Wilson, 2008) and the comprehension of inflected verbs or minimal phrases differentially activate these regions, compared to nouns and nominal phrases (Gennari et al., 2007; Tyler et al., 2008). It might thus be suggested that the computation of semantic roles in a sentence or verb phrase (who is doing what to whom) or of verb morphology (or simply uninflected verbs) are intricately linked to the activation of event representations in the posterior temporal lobe. However, this claim is difficult to reconcile with our results for object static sentences, which clearly involve inflected verbs and arguably have semantic roles like other sentences but do not elicit strong activations in posterior temporal cortex. It appears more likely therefore that subtle event structure properties such as intentionality are responsible for these activations, even though most common verbs will differentially activate this region.

In sum, the evidence suggests that the representations entertained during language comprehension are experience-dependent, as they have been formed through learning from perceptual input. But they need not require the activation of perceptually specific motion representations as reflected on area V5, or the activation of modality-specific motion representations. Instead, these representations appear one-step removed from experience and contain schematic properties of actions and events that include animacy and intentionality information, as suggested by our results and previous perceptual studies. These representations appear organized around modality-specific representations (hence, their proximity to visual cortex) and are triggered by either vision or language, but they are abstracted, generalized or schematized over time through learning, as argued by weak embodiment views, and may act as convergence zone for multiple sources of information, including top-down influences. Nevertheless, constraining contexts or task demands, such as the presence of visual stimulation, may cause V5 to cooperate with their neighboring temporal regions during language comprehension, thus resulting in the engagement of perceptually specific representations. This is expected in a flexible and efficient system that is capable of learning and generalizing from past experience as well as of recruiting modality-specific information, if so required.

## Acknowledgments

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bandl.2013.01.008.

## References

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition, 93*(2), B79–B87.
Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111*(1), 55–71.

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 18*(5–6), 513–562.
Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 84–91.
Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology, 15*(2), 145–153.
Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing stream for manipulable objects and human movements. *Neuron, 34*, 1–20.
Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). FMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience, 15*(7), 991–1001.
Bedny, M., & Caramazza, A. (2011). Perception, action, and word meanings in the human brain: The case from action verbs. *Annals of the New York Academy of Sciences, 1224*, 81–95.
Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., & Saxe, R. (2008). Concepts are more than percepts: The case of action verbs. *Journal of Neuroscience, 28*(44), 11347–11353.
Bedny, M., Caramazza, A., Pascual-Leone, A., & Saxe, R. (2012). Typical neural representations of action verbs develop without vision. *Cerebral Cortex, 22*(2), 286–293.
Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536.
Blakemore, S. J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Segebarth, C., & Decety, J. (2003). The detection of contingency and animacy from simple animations in the human brain. *Cerebral Cortex, 13*, 837–844.
Bornkessel, I., Zysset, S., Friederici, A. D., von Cramon, D. Y., & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *NeuroImage, 26*(1), 221–233.
Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport, 10*, 2619–2623.
Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10*, 649–657.
Chen, E., Widick, P., & Chatterjee, A. (2008). Functional–anatomical organization of predicate metaphor processing. *Brain and Language, 107*, 194–201.
Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence. *Neural Computation, 1*(1), 123–132.
Dumoulin, S. O., Bittar, R. G., Kabani, N. J., Baker, C. L., Jr., Le Goualher, G., Bruce Pike, G., et al. (2000). A new anatomical landmark for reliable identification of human area V5/MT: A quantitative analysis of sulcal patterning. *Cerebral cortex, 10*(5), 454–463.
Friston, K. J. (2012). Ten ironic rules for non-statistical reviewers. *Neuroimage, 61*(4), 1300–1310.
Gennari, S. P. (2012). Representing motion in language comprehension: Lessons from neuroimaging. *Language and Linguistics Compass, 6*(2), 67–84.
Gennari, S. P., MacDonald, M. C., Postle, B. R., & Seidenberg, M. S. (2007). Context-dependent interpretation of words: Evidence for interactive neural processes. *Neuroimage, 35*(3), 1278–1286.
Grewe, T., Bornkessel-Schlesewsky, I., Zysset, S., Wiese, R., von Cramon, D. Y., & Schlesewsky, M. (2007). The role of the posterior superior temporal sulcus in the processing of unmarked transitivity. *NeuroImage, 35*(1), 343–352.
Hauk, O., Johnsrude, I., & Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*(2), 301–307.
Hsu, N. S., Kraemer, D. J. M., Oliver, R. T., Schlichting, M. L., & Thompson-Schill, S. L. (2011). Color, context, and cognitive style: Variations in color knowledge retrieval as a function of task and subject variables. *Journal of Cognitive Neuroscience, 23*(9), 2544–2557.
Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17*(2), 825–841.
Juch, H., Zimine, I., Seghier, M., Lazeyras, F., & Fasel, J. H. D. (2005). Anatomical variability of the lateral frontal lobe surface: Implication for intersubject variability in language neuroimage. *Neuroimage, 24*, 504–514.
Kable, J. W., Lease-Spellmeyer, J., & Chatterjee, A. (2002). Neural substrates of action event knowledge. *Journal of Cognitive Neuroscience, 14*(5), 795–805.
Kable, J. W., Kan, I. P., Wilson, A., Thompson-Schill, S. L., & Chatterjee, A. (2005). Conceptual representations of action in the lateral temporal cortex. *Journal of Cognitive Neuroscience, 17*(12), 1855–1870.
Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., et al. (2005). Perception of motion affects language processing. *Cognition, 94*, B79–B89.
Kellenbach, M. L., Brett, M., & Patterson, K. (2003). Actions speak louder than functions: The importance of manipulability and action in tool representation. *Journal of Cognitive Neuroscience, 15*(1), 30–45.
Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience, 28*(47), 12224–12230.
Kiefer, M., Trumpp, N., Herrnberger, B., Sim, E.-J., Hoenig, K., & Pulvermuller, F. (2012). Dissociating the representation of action- and sound-related concepts in middle temporal cortex. *Brain and Language, 122*, 120–125.

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language, 57*(4), 519–543.

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience, 12*(1), 48–55.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence-intervals in within-subject designs. *Psychonomic Bulletin and Review, 1*, 476–490.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology, 11*(2), 194–201.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex, 48*(7), 788–804.

Morito, Y., Tanabe, H. C., Kochiyama, T., & Sadato, N. (2009). Neural representation of animacy in the early visual areas: A functional MRI study. *Brain Research Bulletin, 79*(5), 271–280.

Noppeney, U., Josephs, O., Kiebel, S., Friston, K. J., & Price, C. J. (2005). Action selectivity in parietal and temporal cortex. *Cognitive Brain Research, 25*(3), 641–649.

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron, 18*(4), 591–598.

Peelen, M. V., Romagno, D., & Caramazza, A. (2012). Independent representations of verbs and actions in left lateral temporal cortex. *Journal of Cognitive Neuroscience, 24*(10), 2096–2107.

Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences, 358*(1431), 435–445.

Pulvermuller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience, 6*(7), 576–582.

Pulvermuller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience, 17*(6), 884–892.

Revill, K. P., Aslin, R. N., Tanenhaus, M. K., & Bavelier, D. (2008). Neural correlates of partial lexical activation. *Proceedings of the National Academy of Sciences of the United States of America, 105*(35), 13111–13115.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition, 76*(3), 269–295.

Rodríguez-Ferreiro, J., Gennari, S. P., Davies, R., & Cuetos, F. (2011). Neural correlates of abstract verb processing. *Journal of Cognitive Neuroscience, 23*(1), 106–118.

Rueschemeyer, S., Glenberg, A. M., Kaschak, M. P., Mueller, K., & Friederici, A. D. (2010). Top-down and bottom-up contributions to understading sentences describing objects in motions. *Frontiers in Psychology, 1*, 1–11.

Saygin, A. P., McCullough, S., Morana, A., & Emmory, K. (2010). Modulation of BOLD response in motion-sensitive lateral temporal cortex by real and fictive motion sentences. *Journal of Cognitive Neuroscience, 22*(11), 2480–2490.

Schultz, J., Friston, K. J., O'Doherty, J., Wolpert, D. M., & Frith, C. D. (2005a). Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron, 45*(4), 625–635.

Senior, C., Barnes, J., Giampietro, V., Simmons, A., Bullmore, E. T., Brammer, M., et al. (2000). The functional neuroanatomy of implicit-motion perception or representational momentum. *Current Biology, 10*(1), 16–22.

Servos, P., Osu, R., Santi, A., & Kawato, M. (2002). The neural substrates of biological motion perception: an fMRI study. *Cerebral Cortex, 12*(7), 772–782.

Schultz, J., Firston, K. J., O'Doherty, J., Wolpert, D. M., & Frith, C. D. (2005b). Activation of posterior superior temporal sulcis parallels parameter inducing the percept of animacy. *Neuron, 45*, 625–635.

Swallow, M. K., Braver, T. S., Snyder, A. Z., Speer, N. K., & Zacks, J. M. (2003). Reliability of functional localization using fMRI. *Neuroimage, 20*, 1561–1577.

Thompson, J. C., Clarke, M., Stewart, T., & Puce, A. (2005). Configural processing of biological motion in human superior temporal sulcus. *Journal of Neuroscience, 25*(39), 9059–9066.

Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., et al. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience, 15*(4), 3215–3230.

Tranel, D., Martin, C., Damasio, H., Grabowski, T. J., & Hichwa, R. (2005). Effects of noun-verb homonymy on the neural correlates of naming concrete entities and actions. *Brain & Language, 92*(3), 288–299.

Tyler, L. K., Bright, P., Fletcher, P., & Stamatakis, E. A. (2004). Neural processing of nouns and verbs: the role of inflexional morphology. *Neuropsychologia, 42*, 512–523.

Tyler, L. K., & Marslen-Wilson, W. (2008). Fronto-temporal brain systems supporting spoken language comprehension. *Philosophical Transactions of the Royal Society of London – Series B: Biological Sciences, 363*(1493), 1037–1054.

Tyler, L. K., Randall, B., & Stamatakis, E. A. (2008). Cortical differentiation for nouns and verbs depends on grammatical markers. *Journal of Cognitive Neuroscience, 20*(8), 1381–1389.

Tyler, L. K., Stamatakis, E. A., Dick, E., Bright, P., Fletcher, P., & Moss, H. (2003). Objects and their actions: Evidence for a neurally distributed semantic system. *NeuroImage, 18*(2), 542–557.

Wallentin, M., Lund, T. E., Ostergaard, S., Ostergaard, L., & Roepstorff, A. (2005). Motion verb sentences activate left posterior middle temporal cortex despite static context. *Neuroreport, 16*(6), 649–652.

Wallentin, M., Nielsen, A. H., Vuust, P., Dohn, A., Roepstorff, A., & Lund, T. E. (2011). BOLD response to motion verbs in left posterior middle temporal gyrus during story comprehension. *Brain and Language, 119*, 221–225.

Whitney, C., Jefferies, E., & Kircher, T. (2011). Heterogeneity of the left temporal lobe in semantic representation and control: Priming multiple versus single meanings of ambiguous words. *Cerebral Cortex, 21*(4), 831–844.

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2011). The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cerebral Cortex, 21*(5), 1066–1075.

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2012). Executive semantic processing is underpinned by a large-scale neural network: Revealing the contribution of left prefrontal, posterior temporal, and parietal cortex to controlled retrieval and selection using TMS. *Journal of Cognitive Neuroscience, 24*(1), 133–147.

Woolrich, M. W., Ripley, B. D., Brady, J. M., & Smith, S. M. T. (2001). Temporal autocorrelation in univariate linear modelling of FMRI data. *Neuroimage, 14*(6), 1370–1386.

Worsley, K. J. (2001). Statistical analysis of activation images. In P. Jezzard, P. M. Matthews, & S. M. Smith (Eds.), *Functional MRI: An introduction to methods*. Oxford: Oxford University Press.

Zihl, J., von Cramon, D., Mai, N., & Schmid, C. (1991). Disturbance of movement vision after bilateral posterior brain damage. Further evidence and follow up observations. *Brain, 114*(Pt 5), 2235–2252.