

# Storytelling Is Intrinsically Mentalistic: A Functional Magnetic Resonance Imaging Study of Narrative Production across Modalities

Ye Yuan, Judy Major-Girardin, and Steven Brown

## Abstract

■ People utilize multiple expressive modalities for communicating narrative ideas about past events. The three major ones are speech, pantomime, and drawing. The current study used functional magnetic resonance imaging to identify common brain areas that mediate narrative communication across these three sensorimotor mechanisms. In the scanner, participants were presented with short narrative prompts akin to newspaper headlines (e.g., “Surgeon finds scissors inside of patient”). The task was to generate a representation of the event, either by describing it verbally through speech, by pantomiming it gesturally, or by drawing it on a tablet. In a control condition designed to re-

move sensorimotor activations, participants described the spatial properties of individual objects (e.g., “binoculars”). Each of the three modality-specific subtractions produced similar results, with activations in key components of the mentalizing network, including the TPJ, posterior STS, and posterior cingulate cortex. Conjunction analysis revealed that these areas constitute a cross-modal “narrative hub” that transcends the three modalities of communication. The involvement of these areas in narrative production suggests that people adopt an intrinsically mentalistic and character-oriented perspective when engaging in storytelling, whether using speech, pantomime, or drawing. ■

## INTRODUCTION

Theories of language origin can be divided into “vocal” and “gestural” models (McGinn, 2015; Arbib, 2012; Armstrong & Wilcox, 2007; MacNeilage & Davis, 2005; Corballis, 2002). Gestural models posit that manually produced symbols evolved earlier than those produced vocally and that speech was a replacement for a preestablished symbolic system that was mediated by gesture alone. Importantly, the kind of gesturing that gestural models allude to is “pantomime” or iconic gesturing. Iconic gesturing through pantomime is thought to have predated symbolic gesturing, passing through an intermediate stage that Arbib (2012) refers to as “proto-symbol.”

From a neuroscientific perspective, these theories of language origin establish a fundamental contrast between two different sensorimotor routes for the conveyance of language, namely, the audiovocal route for speech and the visuo-manual route for pantomime. Language is an inherently multimodal phenomenon, not least through the gesturing that accompanies speaking (Beattie, 2016; Kendon, 2015; McNeill, 2005). Humans have yet a third means of conveying semantic ideas, and that is through the generation of images, as occurs through drawing and writing (Elkins, 2001). We have argued elsewhere that the capacity for drawing is an evolutionary offshoot of

the system for producing iconic gestures such as pantomimes (Yuan & Brown, 2014). Drawing is essentially a tool-use gesture that “leaves a trail behind” in the form of a resulting image. Overall, speech, pantomime, and image generation comprise a “narrative triad,” representing the three major modalities by which humans have evolved to referentially communicate their ideas to one another.

Perhaps, the most important function of language is the communication of narrative, conveying the actions of agents, or “who did what do whom.” Agency is one of the primary elements that is encoded in syntactic structure (Tallerman, 2015). Although word order varies across languages, 96% of languages place the subject (the agent) before the thing that the subject acts upon (Tomlin, 1986). Hence, an “agent first” organization of sentences seems to be an ancestral feature of language grammar (Jackendoff, 1999), and gestural models of language origin highlight this type of sentence organization as well (Armstrong & Wilcox, 2007). Although language is well designed to communicate agency through syntax, it typically does so in a multimodal manner, combining speech and gesture. A basic question for the evolutionary neuroscience of human communication is whether the conveyance of narrative is linked to specific sensorimotor modalities (vocal vs. manual) or whether there are cross-modal narrative areas in the brain that transcend these modalities. This question led us to design an experiment in which we would explore for the first time whether cross-modal

brain areas mediate the communication of narrative ideas using speech, pantomime, and drawing as the triad of production modalities.

Most previous neuroimaging studies of cross-modal communication are perceptual, and we are not aware of production studies that have compared any pair of functions among speech, pantomime, and drawing in healthy adults. Before considering the relevant perceptual studies, we will first examine a handful of studies that have explored the basic network of narrative production, focusing on speech as the output modality. AbdulSabur et al. (2014), in a combined functional magnetic resonance imaging (fMRI) and PET study, had participants learn a series of 12 simple stories, based on a standardized set of three-picture stimuli, and then recount the stories aloud in the scanner when seeing the story's title alone. The control condition was the recitation of standard nursery rhymes. Because this condition involved the production and perception of speech, most of the sensorimotor activations for speech were eliminated in the subtraction of storytelling minus nursery rhymes. In theory, what should be left over are areas involved in the narrative content of the stories. This subtraction revealed areas involved in both language processing and mentalizing (i.e., character processing), including the left inferior frontal gyrus (IFG), dorsomedial pFC, precuneus, superior parietal lobule (SPL), posterior STS (pSTS), cerebellum, and BG. In addition, there was a prominent activation in the cortex of the TPJ, another area that is implicated in both language processing (especially semantics) and mentalizing (Carter & Huettel, 2013) and which, we will argue, is a strong candidate for being a cross-modal hub for narrative processing.

In Hassabis et al. (2014), participants became familiarized before their scan with four characters having contrastive personality traits. In the scanner, participants were required to mentally play out vignettes involving prescribed events happening to the characters in prescribed locations (e.g., the character's drink being spilled while in a bar) and focus on the actions, thoughts, and feelings of the character. Hence, the participant had to mentally simulate a narrative involving the protagonist. Hassabis et al. (2014) observed activations across most of the areas described in the study of AbdulSabur et al. (2014), including the TPJ. Interestingly, nearly identical results were obtained when the participant imagined themselves (rather than a character) in the prescribed scenarios, consistent with the results of Awad, Warren, Scott, Turkheimer, and Wise (2007), in which participants generated self-referential propositional speech ("tell me what you did last weekend"), compared with a baseline condition of counting. The results of these studies suggest that narrative production is not only about recounting a sequence of events but also about conveying embodied episodes in which the perspective of a protagonist is automatically assumed as a default process, as shown by strong activations in the mentalizing

areas like the TPJ, precuneus, posterior cingulate cortex (PCC), and medial pFC. In other words, they suggest a character-driven mechanism of narrative processing in the brain.

Looking now to the multimodal perceptual studies, no neuroimaging study has, to the best of our knowledge, compared images with gestures (although see Wu & Coulson, 2005, for an electroencephalography study). However, several studies have compared speech with gestures. Xu, Gannon, Emmorey, Smith, and Braun (2009) had participants view video clips of an actor performing gestures (pantomimes or emblematic gestures) or listen to an actor speaking words having the same meaning as the observed gestures. A major point of overlap was found not in the TPJ but more ventrally in the pSTS bilaterally. Other studies that have compared speech with gesture have found either similar effects to Xu et al. (2009) in the pSTS alone (Kircher et al., 2009), effects in both the TPJ and pSTS (Redcay, Velnoskey, & Rowe, 2016; Andric et al., 2013), or effects in neither the TPJ nor the pSTS (Straube, Green, Weis, & Kircher, 2012). Cross-modal comparisons between language and images have highlighted similar areas. In an early PET study, Vandenberghe, Price, Wise, Josephs, and Frackowiak (1996) compared the processing of pictures with the processing of single words (visually presented) having the same content as the pictures. They found overlapping activation in the left TPJ (among other areas). Jouen et al. (2015) explored convergent activations related to semantic processing across modalities in an fMRI study in which participants viewed pictures of everyday events or read sentences describing these same types of events. Like Vandenberghe et al. (1996), they found converging activations in the region of the left TPJ bordering on the pSTS. Regev, Honey, and Hasson (2013) found the left TPJ (among other regions) to be an area that was commonly activated across the perception of narratives in spoken and written formats, whereas Wilson, Molnar-Szakacs, and Iacoboni (2008) found the right pSTS to have higher inter-subject correlations for the perception of audiovisual narratives compared with audio-alone narratives. Overall, although the role of the TPJ in language-based narrative is compelling, its importance for processing narrative-based gestures and images is still unclear, with more evidence of convergence being found in the pSTS than more dorsally in the TPJ.

The principal objective of the present fMRI study was to carry out the first three-modality production study of narrative processing, with the aim of identifying a "narrative hub" in the brain. To do this, we had participants read simple headlines (e.g., "Surgeon finds scissors inside of patient") and then depict the narrative described in the headline using either speech (as in a news brief), pantomime (as in the game of charades), or drawing (as in the game of Pictionary), where the latter was done using an MRI-compatible drawing table that allowed participants to see their drawings (Yuan & Brown, 2014). All headlines

described transitive actions carried out by protagonists, in keeping with a view of narrative based on agency. As a way of controlling for sensorimotor differences among the modalities and to hone in on the narrative content per se of the task, we had participants perform a control task in which they were presented with the names of objects (e.g., “binoculars”) and were asked to describe the spatial properties of each object (again either through speech, pantomime, or drawing), while avoiding any mention of the object’s uses or human interactions with it. This permitted a cognitive contrast between “narration” (a recounting of the actions of a protagonist) and “description” (an enumeration of an object’s properties, separate from a person’s interaction with it). We performed the narration-versus-description contrast for each of the three modalities individually. This subtraction permitted us to eliminate the sensorimotor components of the tasks (i.e., audiovocal activations for speech, and visuomotor activations for pantomime and drawing) and thereby isolate components specifically associated with narrative processing of the protagonist’s actions. We then ran a conjunction of the three narration-versus-description subtractions to see if there were any brain areas that were commonly activated across the three narrative modalities of communication, while controlling for sensorimotor differences. On the basis of the literature mentioned above, we predicted that the TPJ and/or the adjacent pSTS would serve such a function. The TPJ in particular is an attractive candidate for a role in cross-modal narrative because it is involved in the processing of language, theory of mind, and agency, hence combining the linguistic and character-related aspects of narrative.

## METHODS

### Stimulus Validation

A set of 60 headline stimuli was devised by the authors. All of them were subject–verb–object declarative statements in the present tense describing narrative events as transitive actions carried out by a protagonist (typically gender-neutral) on some object or person. Examples include “Surgeon finds scissors inside of patient” and “Fisherman rescues boy from freezing lake.” Headlines ranged in length from five to eight words. Word-frequency analysis was performed using the Corpus of Contemporary American English ([corpus.byu.edu/coca](http://corpus.byu.edu/coca)) to ensure that the headlines did not contain words with outlier frequency ratings (i.e., in excess of 1.5 times the interquartile range of the group word frequencies). From a narrative standpoint, the headlines were designed to convey “newsworthy” events that one might find in a newspaper. Half were designed to convey a positive-valenced outcome and half were designed to convey a negative-valenced outcome.

A stimulus validation experiment was carried out using 32 undergraduate students (mean age = 19.8 years, 24 women) to equate the headlines to be used in the

fMRI experiment for level of difficulty across the three modalities of production. Because 60 stimuli could not be accommodated into a 1-hr experiment, the stimuli were randomly divided into four groups of 15 headlines. While in a sound booth, participants were asked to produce a representation of each of the 15 headlines using all three modalities (speaking, pantomiming, and drawing) in a randomized order. Headlines were presented to participants using a laptop computer (HP Pavilion dv5-2050ca) running E-Prime 2.0 Standard (Build 2.0.10.356). Audiovisual behavioral data were recorded using a tripod-mounted video camera (Canon FS200). For each trial, participants were given 4 sec to read the headline and the associated modality of production, followed by a 12-sec planning phase. After a 2-sec “Ready” signal, the word “Begin” indicated the start of a 30-sec production phase. An auditory tone signaled the end of the production phase, and a 5-sec fixation cross was shown before the next trial began. Each of the 45 trials (15 headlines × 3 modalities) lasted 51 sec, for a total session time of just over 38 min.

After the experiment, the participants were asked to rate the headlines that they saw during the experiment for emotional valence (positive, neutral, or negative) and difficulty of production (1 = easy to 3 = difficult) for each headline per modality. Video recordings were used to assess the time required for production. The goal was to assemble a collection of headlines that were not significantly different across modalities in terms of difficulty and that had a completion time of longer than 18 sec to prevent participants from finishing early during the fMRI study (which contained task epochs of 18 sec).

Mean modality difficulty and headline difficulty scores were calculated by collapsing across all modalities and all stimuli, respectively. Emotional valence ratings were tallied to determine how each headline was perceived by participants. A one-way ANOVA was used to test for differences in difficulty scores across speech, pantomime, and drawing for each of the 60 headlines. Post hoc *t* tests were used to examine pairwise differences (i.e., speech vs. mime, mime vs. drawing, and speech vs. drawing). Video recordings of the production times were analyzed in a similar fashion as the questionnaire data using a one-way ANOVA to determine whether there were differences in production times among modalities, with post hoc *t* tests used to examine pairwise differences.

The aim of the validation study was to identify a 24-headline subset of the original 60-headline stimulus set to be used in the fMRI experiment that showed comparable difficulty levels for the three modalities of narrative production. However, not surprisingly, speaking a headline was generally rated by participants to be easier than miming or drawing it. We dealt with this difference in difficulty in two manners. As a first step, we ignored the speech condition and attempted to create a 24-headline set that was equated for perceived difficulty between miming and drawing ( $p > .05$  in the pairwise comparison).

The second step involved introducing a modification of the vocalization style for the speech task. In the validation experiment, participants spoke in a standard manner. However, for the fMRI experiment, we wanted them to speak in a slower and more controlled manner, as has been done in previous vocal studies from our laboratory (Brown, Ngan, & Liotti, 2008). Hence, participants were instructed during a training session on a day before the scanning session to speak with their teeth together so as to reduce jaw movement and to vocalize at a slow pace, roughly two words per second. Compared with the production style of the validation experiment, this should have increased the difficulty level of the speech task. We tested this manipulation on a separate group of undergraduate participants ( $n = 20$ ) and demonstrated that, whereas premanipulation difficulty ratings were comparable between this group of participants and the original validation group, postmanipulation difficulty ratings increased to be similar to the pantomime and drawing difficulty in the validation experiment ( $p > .05$ ).

We would also point out that, whereas pantomime and drawing share the same effector system of the hands, wrists, and arms, speech utilizes a completely different sensorimotor system, composed of the vocal tract and auditory system. In this regard, its presence in the three-way conjunction analysis should bias the results toward erring on the side of false negatives, rather than false positives. Hence, even if the speech task were overall less difficult than the two manual tasks, this would actually minimize a confounding effect of the speech task on the three-way conjunction because it would tend to minimize speech-specific activations compared with the (difficulty-matched) manual tasks.

The control task for the fMRI study consisted of an object description task. Control stimuli were created as single objects that were natural or associative pairs related to concepts present in the headline stimuli. All control objects were inanimate. To minimize the effects of different categorical associations, the control stimuli were chosen to be in the same semantic category as the headline scenarios. Examples included “helicopter” for the headline “Pilot lands plane safely during storm” and “football helmet” for “Quarterback throws long pass to win game.” Note that the control stimuli were words (just as with the headlines), not visual images of objects. Control stimuli were validated by two expert artists with over 10 years of drawing experience. Both artists gave a rating of how well they felt that they could draw the control objects within a time limit of 20 sec using a 5-point Likert-like scale (1 = I did a bad job, 5 = I’m happy with what I drew). All control objects were rated at least 2 by both artists.

## Participants

Twenty-four right-handed individuals who did not participate in the stimulus validation experiment partici-

pated in the fMRI experiment after giving their informed consent (Hamilton Integrated research ethics board, McMaster University). To ensure that the drawing task could be performed in a competent manner, we recruited participants who had a minimum of 2 years of fine arts training. Most of the participants were undergraduate majors in a studio arts program. Two participants were excluded because of head motion, and one participant was excluded for responding to one headline with the incorrect modality, resulting in 21 participants in the analyses (17 women, mean age = 20.4 years). The mean fine arts training of the participants was 5.5 years.

Participants had normal or corrected-to-normal vision (using corrective lenses) and no history of neurological disorders, psychiatric illness, and/or alcohol or substance abuse and were not taking psychotropic medications. They received monetary compensation for their participation. Participants attended a 1-hr training session on a day before the fMRI experiment to become familiarized with the task timing and to learn how to perform all of the tasks in a highly controlled manner so as to minimize head, jaw, and body movements.

## Stimuli

In the MRI scanner, stimuli were presented to participants using a laptop computer (HP Pavilion dv5-2050ca) running E-Prime 2.0 Standard (Build 2.0.10.356). Each headline was paired uniquely to a visual “modality icon” (a voice icon for speech, a hand icon for mime, and a pencil icon for drawing), resulting in 24 headline–modality pairs that did not contain duplicate headlines. No participant produced a given headline with more than one modality (i.e., there was no within-participant repetition), and the full set of headline–modality pairings was achieved in a between-participant manner by creating three stimulus sets across the pool of 21 participants, as produced using a Latin squares approach. Control stimuli (names of objects) were paired to the three modalities using the same approach, but completely independent of the headlines. In other words, although the control objects were generated based on a pairing to the semantic content of the headlines, the control objects were separated from the headlines as follows. The 24 headlines were initially randomly distributed across the four fMRI scans but were then assigned to those scans. Within a scan, they were presented in a random order, but the four scans were presented in a fixed order. Once this assignment of headlines to scans had been made, the control stimuli for the set of headlines for one scan were assigned to another scan, such that the control stimuli derived from a set of headlines were never in the same scan as those headlines themselves. In addition, the three stimulus sets were set up such that a control stimulus was never performed in the same modality as its associated headline. For example, the headline might be mimed in one scan and the control object might be drawn in another scan. As mentioned

above, the performance of all headlines in all three modalities was achieved in a between-participant manner such that no participant ever performed a given headline or object in more than one modality and that no participant ever performed a headline and its associated control object in the same modality or in the same scan.

### fMRI Tasks

During a task epoch, each stimulus (a headline or an object name) and the associated modality icon were displayed for 8 sec, during which time the participants were instructed to plan what they were going to do but to not physically respond. After the planning phase, the screen was replaced by a gray canvas, and participants were given an 18-sec production phase to depict the stimulus item using the assigned modality. There was then a 4-sec “Stop” signal that indicated the end of the production phase, followed by a “Ready” screen for 2 sec as a transition between stimuli. Each task epoch thus lasted 32 sec. For the analysis, the stop and ready periods were eliminated, resulting in 26-sec epochs made up of 8 sec of planning and 18 sec of production. During the fixation trials, a crosshair was displayed for 16 sec, followed by “Ready” for 2 sec, lasting 18 sec, and was analyzed as a single unit. Each of the four scans had a duration of 7 min (420 sec).

Participants performed four fMRI scans composed of both narrative production and object description in all three modalities (speech, mime, and drawing). Each scan consisted of six narrative production trials (two of each modality), six object description trials (two of each modality), and two fixation trials. For the speech modality, participants were instructed to verbally tell a story befitting the given headline or to verbally describe the spatial properties of a given object. For the narrative condition, the instruction during the training session was as follows: “Your task is to tell a story centered on the protagonist of the headline.” As mentioned above, participants were required to use a very slow rate of speech to minimize the difference between the number of elements that could be verbally produced compared with those that could be mimed or drawn. For the mime modality, participants produced bimanual pantomimes to depict a given headline or object. For the drawing modality, participants drew images on an MRI-compatible drawing tablet (Yuan & Brown, 2014; Tam, Churchill, Strother, & Graham, 2010) using their dominant hand. They had full visual feedback of their drawings during the task.

To minimize the narrative content of the object description task, we explicitly instructed participants to focus on the structural properties of the objects and to avoid describing the objects’ uses, because the latter would introduce a human agent into the description. Hence, the contrast between the narrative task and the object task was designed to emphasize the unique presence of a protagonist in the narrative condition.

### Image Acquisition

Functional images sensitive to the BOLD signal were collected with a gradient-echo EPI pulse sequence using standard parameters (repetition time = 2000 msec, echo time = 45 msec, flip angle = 90°, 31 slices per volume, 4-mm slice thickness, no slice gap, matrix size = 64 × 64, field of view = 24 cm, voxel size = 3.75 mm × 3.75 mm × 4 mm), effectively covering the whole brain. Over each scan, 210 volumes of data were collected. Anatomical T<sub>1</sub> images were collected for each participant (3-D fast spoiled gradient echo, inversion recovery prepped, inversion time = 900 msec, echo time = 3.22 msec, flip angle = 9°, receiver bandwidth = 31.25 kHz, NEX = 1, slice thickness = 1 mm, slice gap = 0 mm, field of view = 24 cm, slices = 164, matrix size = 512 × 512).

### Data Analysis

Functional image analyses were conducted using Brain-Voyager QX (Version 2.8.0, Brain Innovation). Images were reconstructed offline, and the scan series was realigned and motion corrected. During the preprocessing stage, a temporal high-pass filter was applied at a frequency of 0.0078 Hz, or two cycles per scan, using the general linear model Fourier algorithm. Three-dimensional spatial smoothing was performed using a Gaussian filter with an FWHM kernel size of 4 mm. After realignment, each functional scan was normalized to the Talairach template (Talairach & Tournoux, 1988). The BOLD response was modeled as the convolution of a 26-sec (task) or 18-sec (fixation) boxcar with a synthetic hemodynamic response function composed of two gamma functions. The six head motion parameters were included as nuisance regressors in the analysis. As mentioned previously, two participants were excluded because of head motion in excess of 1.5 mm of translation or 1.5° of rotation in one or more dimensions, and one participant was excluded for responding to a headline in an incorrect modality. For the 21 participants included in the study, the mean translational head movement was below 1 mm in each dimension, and the mean rotational head movement was below 1° in each dimension. Each participant’s data were processed using a fixed effects analysis, corrected for multiple comparisons using a Bonferroni correction at a threshold of  $p < .05$  for low-level (task vs. fixation) subtractions and false discovery rate (FDR)  $q < 0.05$  for high-level (narrative vs. object) subtractions.

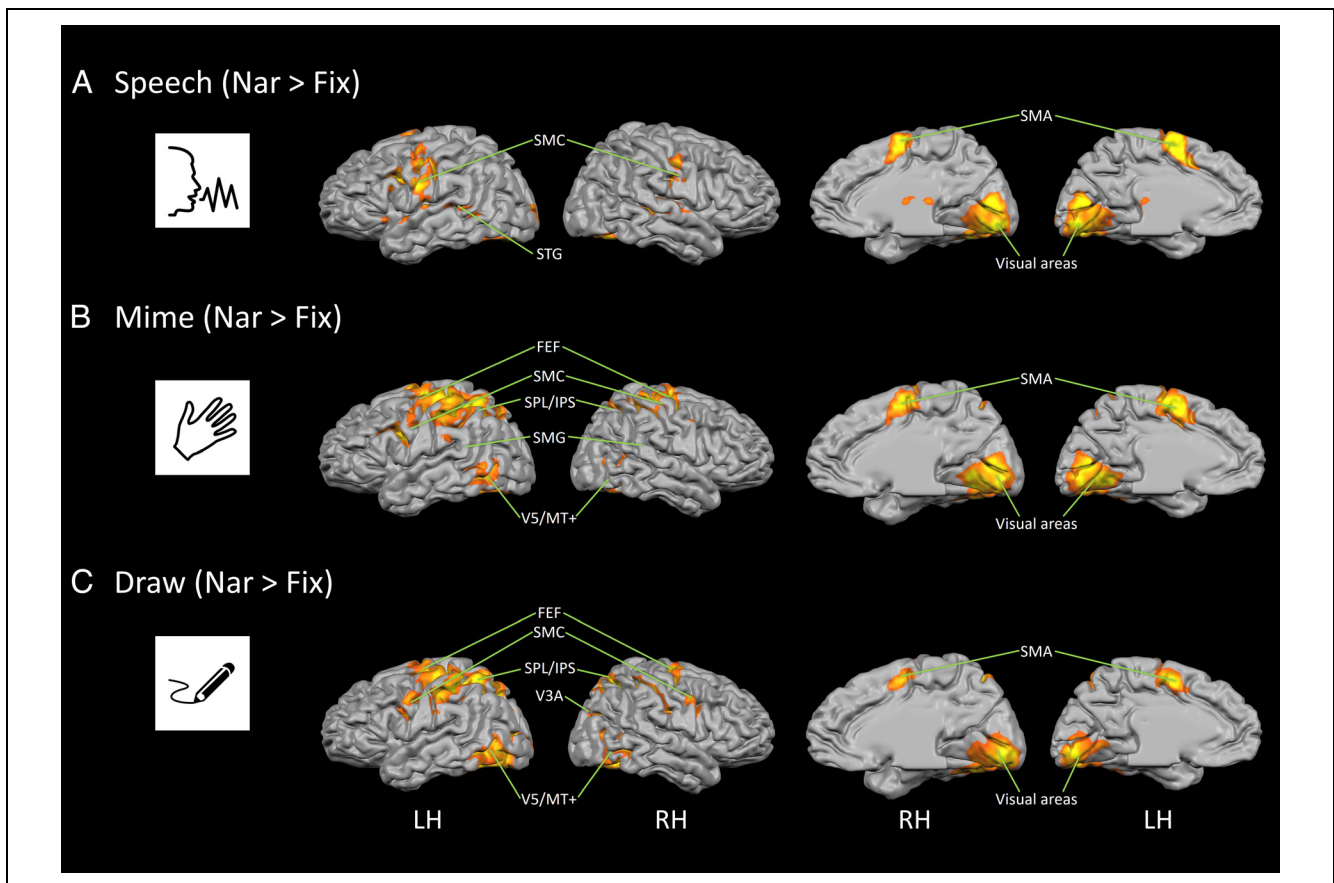
Each individual participant’s results were forwarded into a random effects analysis to produce the group results ( $n = 21$ ). For the low-level subtraction analyses, we contrasted each Task × Modality (Narrative/Object × Speak/Mime/Draw) combination to the fixation baseline condition. This resulted in six subtractions: narrative speech > fixation, object speech > fixation, narrative mime > fixation, object mime > fixation, narrative draw > fixation, and object draw > fixation. We then performed three high-level

intramodal subtractions between narrative production and the object control: (1) narrative speech > object speech, (2) narrative mime > object mime, and (3) narrative draw > object draw. Finally, we performed three pairwise conjunctions between the high-level subtractions for each modality pair as well as the three-way conjunction of interest (“narrative speech > object speech”  $\cap$  “narrative mime > object mime”  $\cap$  “narrative draw > object draw”). All low-level group analyses were thresholded at FDR  $q < 0.001$ ; the higher-order subtractions at FDR  $q < 0.05$ ; and the conjunction analyses at FDR  $q < 0.10$ .

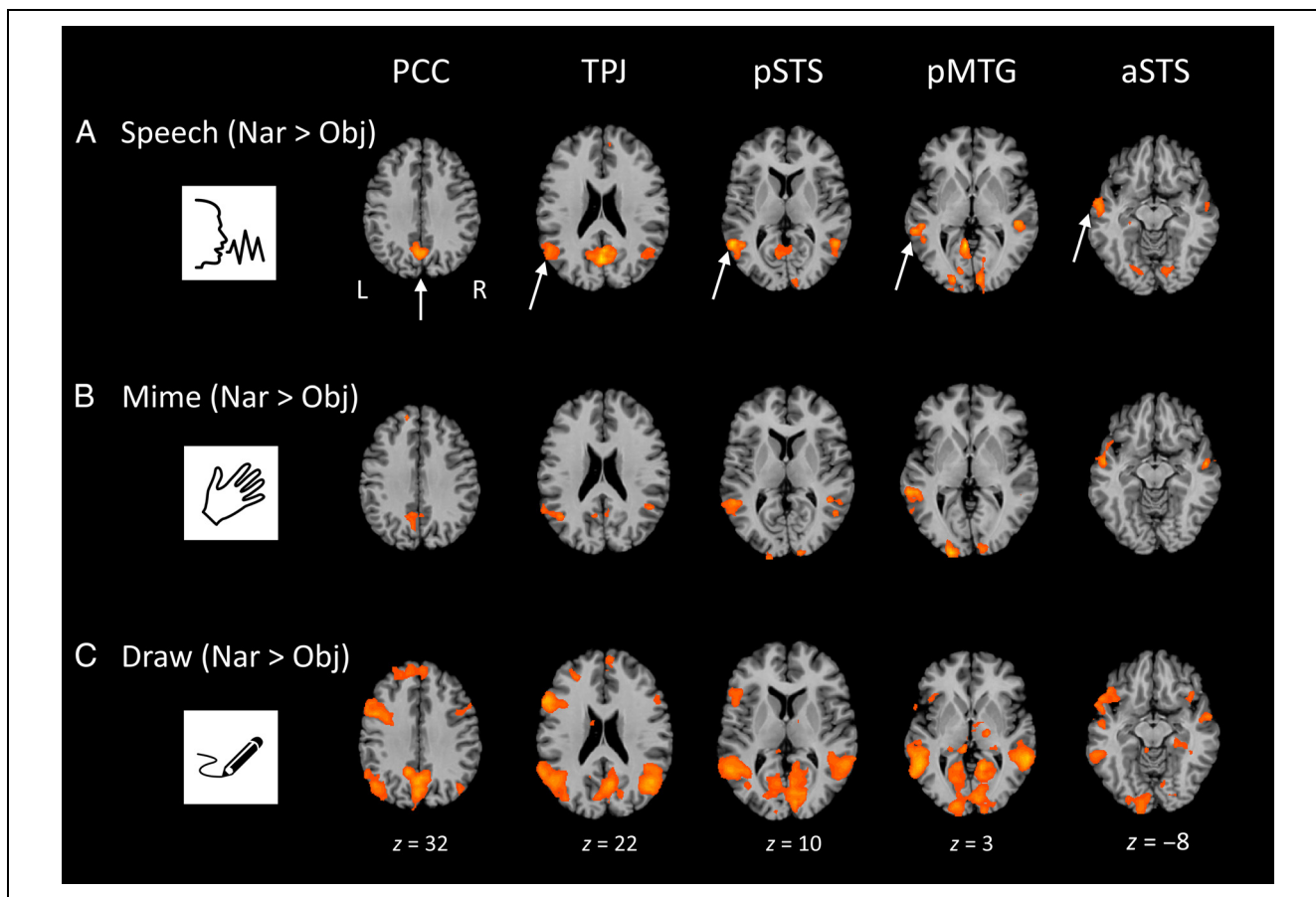
It is important to note that, although the drawing tablet provides information about behavioral performance in the scanner, we were not able to collect behavioral data on speech or mime production in the scanner. Had our focus been on a subtraction analysis, this might have been problematic. However, because our principal goal was to carry out a conjunction analysis, this would emphasize what the three modalities share and would thus offset behavioral differences among modalities with respect to task difficulty.

## RESULTS

Figure 1 shows the brain activations for the narrative > fixation contrast for each modality (FDR corrected,  $q < 0.001$ ; Talairach coordinates in Supplementary Table 1). Prominent activations were found in the primary sensorimotor cortex (SMC) related to the particular effector system employed by each modality: For speech, activity was found ventrally in the orofacial precentral gyrus, whereas for both mime and drawing, activity was found more dorsally in the somatotopic representations for the hand, wrist, and arm. The SMA was found to be commonly activated across the three modalities. Pantomime and drawing, but not speech, showed strong activations throughout the posterior parietal cortex associated with visuomotor processing. This included large parts of the SPL and intraparietal sulcus (IPS), most especially in the left hemisphere. In addition, both of the visuomotor tasks showed activation in the visual–motion area V5/MT+, which was not seen in the speech task. Modality-specific activations included the auditory association cortex of the posterior superior temporal gyrus (pSTG)



**Figure 1.** Narrative > fixation subtractions. Brain activations for the narrative (Nar) > fixation (Fix) contrast for (A) speech, (B) mime, and (C) drawing are presented. All results are displayed at FDR  $q < 0.001$  on a template brain showing the left hemisphere (LH) and right hemisphere (RH). The two brains at the left are lateral views, and the two brains at the right are medial views. The icons in each row are the same ones that participants saw during stimulus presentation. SMG = supramarginal gyrus; V5/MT+ = middle temporal complex. Low-level visual processing areas corresponding to V1 and V2 are labeled as “visual areas.”



**Figure 2.** Narrative > object subtractions. Axial views of the narrative (Nar) > object (Obj) high-level contrast for (A) speech, (B) mime, and (C) drawing are displayed at  $FDR q < 0.05$ . The icons in each row are the same ones that participants saw during stimulus presentation. The figure is set up to show activations in the PCC, TPJ, pSTS, pMTG, and aSTS. These structures are identified with an arrow in A only. Note that prefrontal areas are seen in the PCC column for drawing, whereas visual areas are seen posteriorly in the pSTS, pMTG, and aSTS columns. The Talairach  $z$  coordinate is shown below each column of slices. The left side of the slice is the left side of the brain.

for speech, the left inferior parietal lobule (IPL) for miming, and area V3A for drawing. Lower-level visual areas were seen in all tasks, in part driven by the presentation of the visual prompt during the planning phase (either a headline or an object–word), although the additional visual activity that occurred during task production for pantomime and drawing worked to make the visual activations much stronger than those for speech, including activations occurring bilaterally in the fusiform gyrus.

To isolate the brain areas specific to narrative generation, we performed the high-level narrative > object subtraction for each modality (Figure 2;  $FDR q < 0.05$ ; Talairach coordinates in Table 1). Virtually all of the sensorimotor areas seen in Figure 1—with the exception of low-level visual areas in the lingual gyrus—were eliminated in this subtraction for each modality, suggesting that the control condition was well matched to the narrative condition for these features. What was seen instead were areas associated with mentalizing, social cognition, semantics, and discourse processing. Strikingly similar patterns were seen for all three modalities. The most common areas across the three were the pSTS and TPJ

bilaterally and the PCC at the midline. The posterior middle temporal gyrus (pMTG), a semantic processing area, was present bilaterally for speech and drawing, but only in the left hemisphere for mime. Finally, the anterior STS (aSTS) that is associated with discourse level processing was present bilaterally for all three conditions. Drawing uniquely showed additional activity in the left premotor cortex and dorsomedial pFC.

To look for cross-modal narrative areas in the brain, we ran conjunction analyses for the narrative > object subtractions just described. This included the three pairwise conjunctions and the single three-way conjunction. The results are shown in Figure 3 at a reduced threshold compared with those shown in Figure 2 ( $FDR q < 0.10$ , with Talairach coordinates reported in Table 2 for the three-way conjunction). All pairwise conjunctions showed bilateral activity in the TPJ and pSTS as well as in the PCC at the midline. These areas were also present in the three-way conjunction, although with left-hemisphere dominance. Bilateral aSTS activity was seen in all of the pairwise conjunctions, but only in the left hemisphere in the three-way conjunction.

**Table 1.** Talairach Coordinates for the Narrative > Object Subtractions across Modalities

	<i>Speech (Nar &gt; Obj)</i>				<i>Mime (Nar &gt; Obj)</i>				<i>Draw (Nar &gt; Obj)</i>			
	<i>x</i>	<i>y</i>	<i>z</i>	<i>t</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>t</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>t</i>
<i>Left hemisphere</i>												
pMTG (BA 22/21)	-48	-34	1	6.15	-57	-37	4	6.37	-51	-40	1	8.46
Superior frontal gyrus (BA 8/9)					-12	50	34	6.19	-6	38	46	7.36
									-48	8	37	6.60
TPJ (BA 22/39)	-51	-55	19	7.16					-57	-49	19	6.18
	-48	-61	28	6.36	-39	-58	25	5.76	-45	-64	28	6.56
pSTS (BA 22/21)					-45	-46	13	6.44	-51	-52	7	6.97
PCC (BA 7/31)	-3	-61	34	6.29	-12	-55	37	5.25	-3	-64	37	7.10
	-9	-58	25	6.22					-12	-52	31	6.06
	-6	-58	46	5.72					0	-76	34	5.92
aSTS (BA 21)	-60	-4	-8	6.09	-51	-7	-8	6.20				
IFG (BA 47)	-57	11	-2	5.59	-42	23	-11	6.45	-42	17	-2	6.09
Premotor cortex (BA 6)									-33	-4	40	7.30
									-39	-4	52	6.91
IFG (BA 45)									-54	20	22	7.05
Lingual gyrus (BA 17)									-18	-94	1	7.04
Superior frontal gyrus (BA 6)									-12	-1	64	5.97
Temporal pole (BA 38)									-51	17	-8	5.81
Thalamus (pulvinar)									-9	-34	7	5.64
<i>Right hemisphere</i>												
PCC (BA 31)	6	-64	28	8.21					6	-67	28	5.66
TPJ (BA 39/22)	45	-61	22	6.26	48	-55	19	6.00	45	-58	22	6.54
									48	-67	25	5.89
pSTS (BA 22)	51	-52	16	7.17	42	-43	13	5.78	57	-46	13	6.01
pMTG (BA 21)	48	-31	-2	6.44					51	-25	1	6.36
	63	-31	-2	5.77					60	-40	4	6.73
aSTS (BA 21)	54	-13	-5	6.35	54	-7	-8	6.41	54	-1	-5	5.59
IFG (BA 47)					51	23	-11	6.64				
Ventral PCC (BA 31)									6	-64	16	6.87
Lingual gyrus (BA 17)					15	-94	10	5.82	9	-82	10	6.12
Dorsomedial pFC (BA 9)									6	50	37	5.75

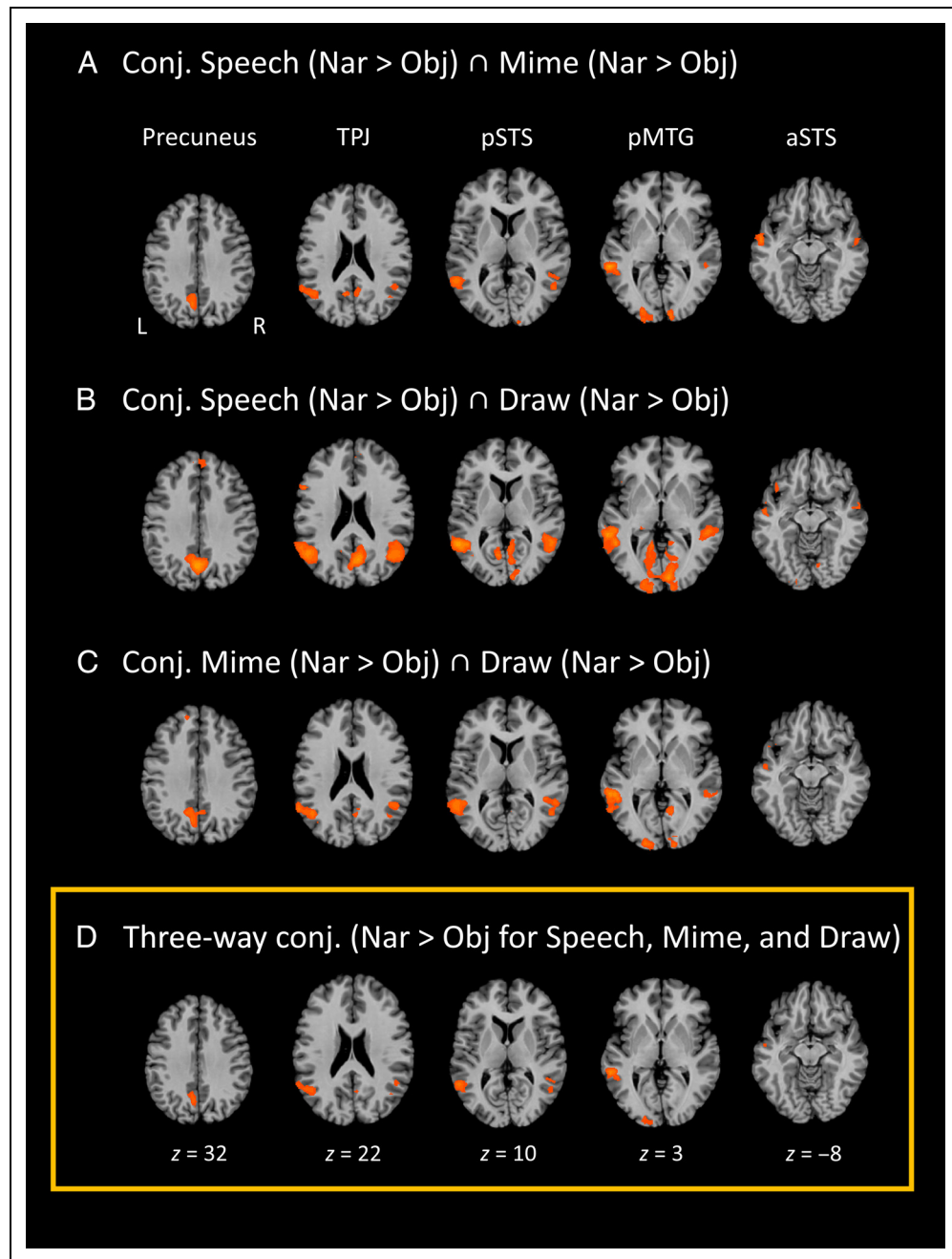
Talairach coordinates are presented for the peak activations for the narrative > object subtraction for speech, mime, and drawing. Brodmann's areas (BAs) are indicated in parentheses. The columns labeled as *x*, *y*, and *z* contain the Talairach coordinates for the peak of each cluster reaching significance. The *t* value is the maximal value for that cluster. To keep the table to a manageable size, only peaks with a *t* value > 5.5 are shown. Nar = narrative; Obj = object.

Finally, we decided to look at the results of the three-way conjunction separately for the planning and production phases of the task, as shown in Figure 4 (with Talairach coordinates in Table 2). The mentalizing areas

that were present when both phases were combined were mainly activated during the production phase alone. This included the TPJ, PCC, and pSTS. The planning phase was dominated by visual activations related to stimulus



**Figure 3.** Conjunction analyses. Axial views of the conjunction results from the narrative (Nar) > object (Obj) contrasts are presented. The first three panels show pairwise conjunctions for (A) speech and mime, (B) speech and drawing, and (C) mime and drawing, respectively. The bottom panel shows the three-way conjunction of speech, mime, and drawing. The figure is set up to show activations in the PCC, TPJ, pSTS, pMTG, and aSTS. Note that visual areas are seen posteriorly in the pMTG column. Results are displayed at FDR  $q < 0.10$ . The Talairach  $z$  coordinate is shown below each column of slices. The left side of the slice is the left side of the brain.



processing (including the lateral geniculate nucleus) as well as the left pSTS, left pMTG, and left aSTS. Figure 4 shows an approximate dorsoventral distinction, such that the planning phase mainly activated ventral areas and the production phase activated more dorsal areas.

## DISCUSSION

The principal objective of this fMRI study was to carry out the first three-modality production study of narrative processing, with the aim of identifying a narrative hub in the brain. We examined this by looking at the contrast between a narration condition and an object-description condition

and then conjoining this contrast across the three major modalities of narrative expression: speech (language), pantomime, and drawing. In keeping with our predictions based on the narrative literature for speech, we found an effect in the TPJ as well as in the pSTS and PCC. These areas are strongly associated with “character” processing, as related to both mentalizing and emotional expression (including a person’s facial expression, vocal prosody, and expressive body motion). We will argue below that a potentially unifying explanation for our activation profile might relate to the concept of agency.

During both everyday conversation and performances of theatrical works, people tell stories about themselves

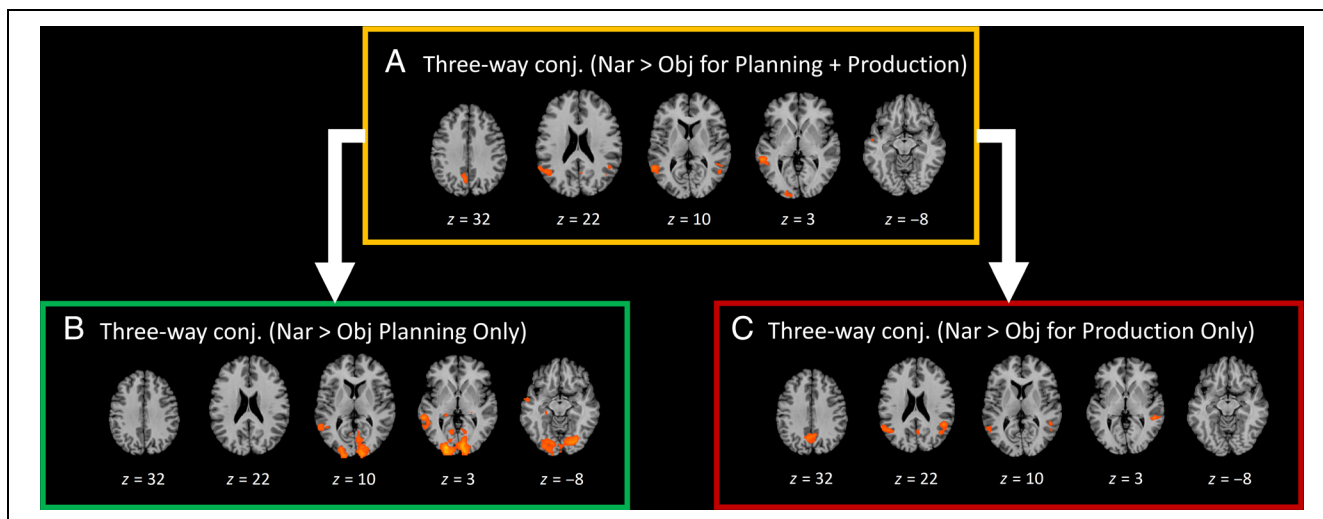
and others through a process of narration. In most stories, there is a central protagonist (be it oneself or a character) who drives the actions of the narrative and who serves as a focal point by which perceivers (listeners or readers) understand the goal structure of the story's actions (Abbott, 2008; Mandler, 1984; Stein & Glen, 1979; Rumelhart, 1975). As a result, people see the central conflict of the story from the protagonist's per-

spective. The narration condition of our experiment was designed to tap into protagonist processing. Because objects were part of both the control condition and the narrative condition (the headlines described transitive actions on objects), what was unique to the narrative condition was a protagonist and his or her actions. Hence, the activation results most likely reflect this.

**Table 2.** Talairach Coordinates for the Three-Way Conjunction

	<i>Three-Way Conj. (Plan + Prod)</i>				<i>Three-Way Conj. (Plan Only)</i>				<i>Three-Way Conj. (Prod Only)</i>			
	<i>x</i>	<i>y</i>	<i>z</i>	<i>t</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>t</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>t</i>
<i>Left hemisphere</i>												
pMTG (BA 21)	-57	-37	1	5.56	-57	-37	4	5.48	-51	-31	-2	4.73
aSTS (BA 21)	-54	-4	-11	5.22	-54	-7	-8	4.61				
pSTS (BA 22/39)	-54	-52	13	5.27	-42	-49	13	3.78	-51	-52	13	4.74
TPJ (BA 39)	-42	-55	22	4.76					-48	-58	25	4.94
PCC (BA 7)	-6	-64	34	5.05					-6	-61	34	5.49
IFG (BA 47)	-39	20	-2	4.53								
Premotor cortex (BA 6)					-36	-4	49	4.73				
Dorsomedial pFC (BA 8)					-9	47	40	4.14				
<i>Visual areas</i>												
Lingual gyrus (BA 17)	-18	-91	1	4.88	-18	-88	4	7.40				
Middle occipital gyrus (BA 18)					-18	-97	10	5.50				
Fusiform gyrus (BA 19)					-24	-79	-14	5.42				
Inferior occipital gyrus (BA 18)					-27	-85	-5	4.81				
Lateral geniculate nucleus					-21	-28	-2	4.89				
<i>Right hemisphere</i>												
TPJ (BA 22)									48	-46	22	5.55
pSTS (BA 22)									51	-55	16	5.02
pMTG (BA 22)									51	-37	4	5.24
aSTS (BA 21)									48	-19	-5	4.88
PCC (BA 31/7)									3	-58	37	5.37
Ventral PCC (BA 31)					6	-67	13	4.81				
<i>Visual areas</i>												
Lingual gyrus (BA 17)	12	-97	4	4.04	12	-94	4	6.24				
Fusiform gyrus (BA 19)					27	-70	-11	5.49				
Middle occipital gyrus (BA 18)					21	-85	13	4.54				
Lateral geniculate nucleus					24	-28	-2	4.10				

Talairach coordinates are presented for the three-way conjunction (Conj.) of the narrative > object subtractions for speech, mime, and drawing. They are organized by the combination of planning and production (Plan + Prod), planning only (Plan), and production only (Prod). Brodmann's areas (BAs) are indicated in parentheses. The columns labeled as *x*, *y*, and *z* contain the Talairach coordinates for the peak of each cluster reaching significance. The *t* value is the maximal value for that cluster.



**Figure 4.** Planning and production in the three-way conjunction. Axial views of the three-way conjunction are presented with respect to (A) the planning and production phases combined, (B) the planning phase alone, and (C) the production phase alone. A is a reproduction of Figure 3D. The figure has the identical slice levels as Figures 2 and 3. Results are displayed at FDR  $q < 0.10$ . The Talairach  $z$  coordinate is shown below each slice. The left side of the slice is the left side of the brain.

Two major neural systems have been invoked in understanding human action, including through narratives. The mirror neuron system is a sensorimotor system that is thought to mediate an understanding of the actions of agents (Arbib, 2012). From our standpoint, we can think of this system as being a “gestural” system, because it deals with the visually perceivable motor gestures of an agent. This is in contrast to the mentalizing system, which deals with inferring the unobservable mental states of these agents (Frith & Frith, 2003; Nichols & Stich, 2003). Although our narrative task was, on the surface, quite gestural—the headlines were statements of transitive actions with no explicit mentalistic content—a key question for the interpretation of the imaging results is whether people intrinsically tend to adopt a mentalistic stance when they recount stories about protagonists, even if the task does not require them to infer the mental states of the agents. If so, it would suggest that people approach storytelling less as a recounting of the event sequences that make up the plot (as in studies of “event perception”; Radvansky & Zacks, 2011) than as a connection with the mental states of the characters. In other words, people may carry out storytelling in a character-based, rather than plot-based, manner as their default mode of operation. In fact, even for studies of event perception, the narrative-related activations that we observed are those that are associated with characters, rather than objects, space, or causation in their narrative stimuli (Zacks, Speer, Swallow, & Maley, 2010).

### The Unimodal Analyses: Low- and High-level Subtractions

The low-level subtractions against fixation (Figure 1) revealed mainly sensorimotor areas involved in task performance. On the production side, this included the SMC

and SMA. For the SMC, we observed the expected somatotopic distinction between two general effector systems used for communication: the orofacial motor cortex for the speech task and both the hand and arm motor areas for the two manual tasks. The pantomime and drawing conditions also showed prominent activations in the posterior parietal cortex not seen in the speech condition, including the SPL, IPS, and IPL. Such activations are thought to mediate visuomanual translation during visually guided hand movement tasks and are common areas of activation across studies of both pantomime (Vingerhoets & Clauwaert, 2015) and drawing (Yuan & Brown, 2014, 2015). In addition, perceptual areas that were stimulated by the outcome of the production process included auditory areas for speech and visual areas for miming and drawing. The latter included area V3A that we have described previously as being important for perceiving the image that dynamically accumulates as a result of the process of drawing (Yuan & Brown, 2014; see also Thaler & Goodale, 2011). This activation was considerably weaker in the pantomime condition, in which this type of visual accumulation does not occur. Activations in low-level visual areas were also due to the presence of the stimulus prompts and were therefore present in the speech conditions as well (Figure 1).

Next, the high-level subtractions were designed to eliminate the abovementioned sensorimotor activations and thereby isolate components specifically associated with narrative processing of the protagonist’s actions. This permitted a cognitive contrast between narration (a recounting of the actions of a protagonist) and description (an enumeration of an object’s properties, separate from a person’s interaction with it). The efficacy of the control condition was shown by the fact that most of the sensorimotor areas just described were eliminated

in the narration-versus-description subtraction for each modality (Figure 2), suggesting that sensorimotor activations were well accounted for by the control condition. The major exception consisted of visual processing areas in the lingual and fusiform gyri that were more active for narrative processing than object processing, most especially for drawing. Hence, drawing the complex scenes of the narrative scenarios composed of multiple objects stimulated greater visual processing than creating detailed drawings of single objects.

What was left over in each case were areas associated with the mentalizing and social cognition networks, which will be discussed in the next section on the conjunction analysis. For now, we focus on two additional areas that came up in the high-level subtractions, namely, the aSTS (BA 21/22/38) and the pMTG (BA 21). The aSTS came up in the high-level subtractions bilaterally for all three modalities. This is a part of the brain that, in auditory perception studies, is thought to show a preference for stimuli that are complex and coherent, compared with either elemental stimuli or complex stimuli that are meaningless or incoherent (Scott, Rosen, Lang, & Wise, 2006; Scott, Blank, Rosen, & Wise, 2000). For example, it shows a preference for the perception of sentences, compared with words or phonemes (DeWitt & Rauschecker, 2012). This supports a role of the anterior temporal region in discourse processing, text integration, and the generation of meaning beyond the single-sentence level (Mason & Just, 2006), something that is important for narrative processing (Mar, 2011). In a study by Brown, Martinez, and Parsons (2006) on the vocal generation of completions for both sentence fragments and musical melody fragments, the authors found the aSTS to be activated in both the speech and music conditions. It is interesting that we saw activation of this area for pantomime in this study, because pantomime tends to share with speech its linear, sequential nature, whereas individual drawings are not typically created in a linear and episodic fashion. An exception is the processing of comic strips, whose perception does show activation in the anterior temporal region (Osaka, Yaoi, Minamoto, & Osaka, 2014). This suggests that our artist participants might have developed a linear strategy for drawing the scenarios. Regardless of modality, the linear stringing together of events is a critical part of narrative processing.

The pMTG's connection with narrative relates to its central role in semantic processing. Not only is the pMTG a key semantic area of the brain (Binder, Desai, Graves, & Conant, 2009), but its function is strongly cross-modal, as shown by its activation in semantic tasks using visual words, spoken words, pictures, and gestures (Krieger-Redwood, Teige, Davey, Hymers, & Jefferies, 2015; Visser, Jefferies, Embleton, & Lambon Ralph, 2012; Xu et al., 2009). Hence, the pMTG might function as an amodal semantic area, although some studies link it most strongly to visual processing (Devereux, Clarke, Marouchos, & Tyler, 2013). In Huth, de Heer, Griffiths, Theunissen, and

Gallant's (2016) fine-grained analysis of semantic processing throughout the brain, the pMTG was found to be associated with "numeric, tactile, and visual concepts," whereas the TPJ was found to be associated with "social, emotional, and mental concepts." This suggests that, whereas the TPJ might deal with semantic concepts related to the characters of a story, the pMTG might deal with more general semantic aspects of the story's setting, such as visual features of the scene. Given that the pMTG projects to the TPJ region via the vertical limb of the arcuate fasciculus (Catani, Jones, & ffytche, 2005), this might suggest a sequential model of narrative processing in which "setting" information in the pMTG is transmitted to and combined with character information in the pSTS and TPJ. This sequential interpretation is supported by the conjunction analysis, which showed that (1) the left pMTG was more strongly present during the planning phase than the production phase, (2) the pSTS was present in both phases, and (3) the TPJ (and PCC as well) was present in the production phase alone. Hence, this is suggestive of a ventral-to-dorsal projection of information in the posterior temporal lobe.

### **The Cross-modal Analyses: Two- and Three-Way Conjunctions**

The three unimodal subtractions showed strikingly similar patterns of activation among themselves, focusing mainly on areas of the mentalizing and social cognition networks involved in character processing. There were few additional brain areas in each subtraction beyond those that were common across the three (Figure 2); as mentioned, these included visual areas in addition to some premotor and prefrontal areas for drawing. As a result, the cross-modal conjunctions, whether two- or three-way, showed these same general areas. These included the five major areas of TPJ, pSTS, PCC, pMTG, and aSTS. These areas were not seen for the most part in the narrative condition against fixation but mainly came up in the high-level contrasts. ROI analyses (not shown) demonstrated that this occurred either through a reduction in the relative level of deactivation between object and narrative processing (PCC) or through an increase in the relative level of activation (e.g., pSTS).

The only one of the five cross-modal areas with a clear association with sensory processing is the pSTS, although it too has a connection with "character" processing to the extent that it is involved in multimodally perceiving the speech prosody, facial expression, and body movement of people as part of the process of social cognition (Biau, Morís Fernández, Holle, Avila, & Soto-Faraco, 2016; Deen, Koldewyn, Kanwisher, & Saxe, 2015; Kreifelts, Ethofer, Shiozawa, Grodd, & Wildgruber, 2009; Campanella & Belin, 2007). As mentioned in the Introduction, the pSTS is a prominent area of overlapping activity in cross-modal studies of language and narrative processing for speech, gesture, and visual images.

Although the pSTS is typically seen in studies of perception, we demonstrate its role in production as well, as seen in previous studies of narrative production (AbdulSabur et al., 2014). Adding to the pSTS's dual role in production and perception, the conjunction analysis revealed that the pSTS was the one cross-modal area that was common to planning and production as well.

### **A Cross-modal Narrative Hub?**

Humans possess a “narrative triad” of three major modalities for conveying narratives, each of which is highlighted in one of the general branches of the arts: speech/language in theater and literature, pantomime in mime theater and narrative forms of dance, and image generation in drawing, painting, sculpting, and related forms of graphic art. Our experiment attempted to factor out the modality-specific components of narrative to identify an area or a set of areas that would be common to the multiple forms of narrative expression and would thus serve as an amodal conceptual area for narrative generation. One hypothesis that we had in mind for this was that a cross-modal production area should also be a region of production/perception overlap, because such an overlap would be one indicator of the pleiotropy of the brain area. Although we did not examine narrative perception per se in our experiment (perception was an integral component of all the task conditions), both Silbert, Honey, Simony, Poeppel, and Hasson (2014) and AbdulSabur et al. (2014) directly compared narrative production with perception and found an overlap in a large number of areas. Among them were all of the areas seen in our conjunction analyses, including the TPJ, pSTS, PCC, pMTG, and aSTS. Hence, this satisfies our criterion that these areas be multifunctional, not only cross-modally (as described in the Introduction) but also sensorimotorically. As mentioned above, the pSTS was active during both planning and production in the current experiment, whereas the TPJ and PCC were active principally during the production of narratives.

The activations obtained in both the high-level subtractions and conjunctions suggest that narrative production is more associated with character than plot, despite the field of literary studies prioritizing plot over character since the time of Aristotle. The activations showed an orientation toward encoding features of the protagonist. The TPJ and PCC are components of the classic mentalizing network involved in making inferences about the beliefs, desires, and emotions of other people as well as oneself (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; Denny, Kober, Wager, & Ochsner, 2012; Lombardo et al., 2010; Frith & Frith, 2003, 2006; Gallagher & Frith, 2003). This network is typically probed in neuroimaging studies with tasks that have participants explicitly think about the beliefs and emotions of others, although evidence suggests that the same network is involved in implicit mentalizing as well (Kestemont, Vandekerckhove,

Ma, Van Hoeck, & Van Overwalle, 2013; Van Overwalle & Vandekerckhove, 2013). In either case, the mentalizing network is virtually always seen in perceptual tasks rather than production tasks.

As mentioned above, there was nothing overtly mentalistic about our narrative condition. Participants were asked to recount short narratives about transitive actions conveyed in the headlines. Nothing about the task required them to mentalize about the protagonist. Given that the mirror system emphasizes observable gestures and that the mentalizing system emphasizes unobservable mental states, our task would seem to better fit the profile of the mirror system, because it was oriented toward describing the observable behaviors (gestures) of an agent as well as event sequencing. A gestural emphasis is seen in studies of pantomime production, where people have to act out simple actions or represent objects (Schippers, Gazzola, Goebel, & Keysers, 2009). One could argue that, by removing the neural encoding of the object through our control condition, what would be left over would be the action itself, hence perhaps reflected in the activation of premotor areas or related areas involved in motor sequencing, such as IFG and/or BG (Shmuelof & Krakauer, 2011). However, we did not see that pattern. It appears that participants, instead of focusing on the actions per se, focused on the protagonist and his or her underlying mental states. Perhaps the activation most indicative of a gestural interpretation was the pSTS. However, even that area reflects the social aspects of communicative expression, rather than motoric parameters related to, say, grasping an object (as the surgeon might grasp the scissors found inside the patient in one of our headlines), which is the classic stimulus for the mirror system.

One possible interpretation of the results is that people assume a mentalistic stance as their default mode of processing stories, oriented toward the characters in the stories. Because we did not manipulate any aspect of mentalizing in our narrative condition, we had no control over whether participants did or did not process the headlines mentalistically when performing them. Although we can easily imagine that participants could have engaged in this mode of processing for speech, we find it unlikely that they would have done so for drawing, where participants simply drew a scenic description of the headline, for example, a patient lying on an operating table and a surgeon next to him finding a pair of scissors inside him. Regarding pantomime, there are two distinct modes of performing a pantomime (Suddendorf, Fletcher-Flinn, & Johnston, 1999; O'Reilly, 1995; Boyatzis & Watson, 1993): egocentrically as an open-handed gesture (as in miming a tennis serve with an imaginary racquet in one hand and an imaginary tennis ball in the other) or allocentrically as an act of object substitution (as such when a person mimes “call me” by using their hand to embody a telephone receiver). Although these two modes of pantomime involve strikingly different relationships between the mimer's body

and the mimed object, it does not seem that one format is inherently more mentalistic than the other, although one could argue that the egocentric mode is more self-oriented and that the allocentric mode is more other-oriented.

It is important to state that, because our control task was an object task, we cannot rule out an interpretation of our data based on event processing. Hence, future studies will be required to disentangle character- versus event-based contributions to narrative processing. For example, we could directly manipulate mentalistic processing by the participants in a way that was not done in the current study. We could compare a recounting task (i.e., simply describing what happened without reference to motivation or causation) with a mentalistic task (i.e., describing what happened with reference to the protagonist's goals and the consequences therein). In the study of Schippers et al. (2009), participants either produced or perceived pantomimic gestures with the aim of gauging the relative involvement of the mirror and mentalizing systems. Although components of both systems were involved in perceiving pantomime (including the TPJ), only mirror neuron areas were involved in production. Half of the stimuli in that study were objects, which would have fit into our control condition. The other half were simple actions, for example, "peel fruit." We instructed participants in our study to not mime the uses of objects in the control condition because pilot testing had shown that this reduced activity in mentalizing areas, most likely because of the virtual agent that underlies egocentric pantomiming of actions.

Similar results to ours were obtained in a study of visual narrative perception by Deghani et al. (2017), which looked at story comprehension in speakers of three different languages (English, Farsi, and Mandarin) and which extracted the language-independent brain activations that were common to all three sets of native speakers. Such activations occurred most prominently in the TPJ, PCC, and dorsomedial pFC bilaterally (among other areas). The authors interpreted their results in the context of the default mode network, rather than the mentalizing network, although the two networks are highly overlapping. It is worth noting that, in this study's reading task, each story was preceded by a context slide identifying the protagonist of the story and each story was followed by a question asking about the personal values of the protagonist. As a result, the character-centered processes that we are discussing here were most likely highly engaged in the reading task of Deghani et al. (2017).

Overall, our results suggest that, when people generate narratives, they assume a mentalistic stance that is driven by their psychological conception of the protagonist, rather than a purely gestural approach to the observable event sequence of actions carried out by the protagonist. In doing so, they cognitively prioritize character over plot. Although the mentalizing system is invoked in perceptual studies of narratives (Mar, 2011), what is lacking is a comparable production system for

narratives, not least one from a character-centered perspective. The mirror system is the classic system for aligning perception and production. However, it is principally a gestural system. A theoretical understanding of narrative requires a comparable sensorimotor system for mentalistic production and perception. In fact, we would argue that "narration is the production counterpart to the perceptual process of theory of mind." If theory of mind is the private process of inferring a character's motivations, beliefs, and emotions, narration is the public process of externalizing such motivations, beliefs, and emotions through depictive acts of communication (Clark, 2016). Comparisons between story production and perception (AbdulSabur et al., 2014; Silbert et al., 2014) have demonstrated that all of the areas seen in our high-level subtractions and conjunctions were comparably present in production and perception. Other areas included the IFG, anterior temporopolar cortex, and dorso-lateral pFC.

One concept that has the potential to unify many key aspects of narrative processing is "agency," which refers to the intentionality to act and the sense of voluntary control over one's actions and one's ability to achieve desired outcomes (Haggard, 2017). The vast majority of the psychology literature on agency deals with "self-agency," in other words, perceptions about one's own capacity to act. What is lacking is a concept of "other-agency" that seems to underlie third-person storytelling. Much narrative, whether during conversation or in literature, is about recounting the agency of protagonists. According to both literary theory and psychological models of story grammars, stories are generally about the goal-directed behaviors of protagonists, their efficacy at overcoming obstacles, and their capacity to achieve their goals through intentional actions, often to solve problems (Abbott, 2008; Mandler, 1984; Stein & Glen, 1979; Rumelhart, 1975). Stories are very much outcome driven, based on an arc-like sequence of events (the "dramatic arc") that resembles the psychological progression of problem-solving episodes. Narrative models contain all of the ingredients of models of self-agency but apply them to "others." Hence, storytelling is typically a third-person recounting of the agency of protagonists.

Interestingly, neuroimaging analyses of instrumental agency show an effect in the TPJ. However, the connection is much stronger with "external" agency than self-agency. External agency, in the context of the operant actions looked at in this literature, refers not to social agency in human interactions (as per stories) but rather to the ability to control the outcomes of instrumental actions, such as pushing a button to cause a tone to be emitted. To the best of our knowledge, there is minimal neuroimaging literature exploring social agency between interaction partners, not least the agency of people other than oneself. One function of the TPJ activations seen in our subtractions and conjunctions could be related to the attribution of agency to protagonists, again consistent

with a general orientation toward characters, rather than episodic sequences, in stories. It is interesting to point out that a meta-analytic comparison of other-judgments versus self-judgments in mentalizing paradigms revealed that many brain areas were commonly activated between other- and self-processing, whereas the TPJ bilaterally and the PCC were among a small number of areas that showed a preference for other-processing over self-processing in the direct contrast (Denny et al., 2012). This is again compatible with the notion that the TPJ might be specialized for processing other-agency more than self-agency, as would be important in narrative production and perception. Overall, components of the mentalizing network (TPJ and PCC) in combination with the pSTS seem to constitute a set of hub areas for narrative production, with a special focus on the protagonists, perhaps related to their agency. Their preferential activation during the production phase of our task, rather than the planning phase, supports an interpretation of character-driven narrative generation.

### Evolutionary Implications

Both vocal and gestural models of language attempt to account for the origins of syntax. As mentioned in the Introduction, language grammar seems to have an intrinsically narrative structure to it, being efficient at describing who did what to whom—in other words, agency. Standard subject–verb–object models of syntactic structure (Tallerman, 2015) essentially encapsulate the kinds of transitive actions that we examined in our headlines. A large majority of languages operate on an agent-first basis, putting the actor before either the action or the target of the action. To the extent that agency is one of the most fundamental things that is conveyed in grammars (and which is lacking in so-called proto-languages; Bickerton, 1995), then our results have application to evolutionary models of language. In particular, the imaging results that were obtained in the most purely linguistic condition (speech) were replicated almost identically in the nonlinguistic conditions of pantomime and drawing. This cross-modal similarity suggests that the capacity of syntax to represent agency can be achieved through nonlinguistic means employing essentially the same brain network.

A number of biological theories of language propose that syntax emerged from basic processes of motor sequencing (Arbib, 2012; Fitch, 2011; Jackendoff, 2011). Although this might account for grammar's connection with object-directed actions—in other words, the gestural level of representation—it may not do justice to the sense of agency that is well contained in syntactic structure. Hence, we suggest that another important evolutionary ingredient in the emergence of syntax—beyond the “plot” elements contained in motor sequencing—would be the incorporation of circuits that mediate the sense of agency, not least “other” agency. To be clear, we are not arguing that the TPJ and pSTS are syntax areas. We are simply

suggesting that, whereas circuits in the IFG more typically associated with syntax (Zaccarella & Friederici, 2017) might mediate the gestural level of language, the TPJ might have a stronger connection with agents in the overall scheme of language, discourse, and narrative. Agency can be conveyed linguistically through speech and sign, but it can also be conveyed nonlinguistically through pantomime (iconic gesturing) and drawing.

### Conclusions

In this first three-modality fMRI study of narrative production, we observed results that suggest that people generate stories in an intrinsically mentalistic fashion focused on the protagonist, rather than in a purely gestural manner related to the observable action sequence. The same set of mentalizing and social cognition areas came up with each of the three modalities of production that make up the narrative triad, pointing to a common set of cognitive operations across modalities. These operations are most likely rooted in character processing, as related to a character's intentions, motivations, beliefs, emotions, and actions. Hence, narratives—whether spoken, pantomimed, or drawn—seem to be rooted in the communication of “other-agency.”

### Acknowledgments

We thank Raymond Mar for critical reading of the article and for his helpful suggestions for improvement. This work was supported by a grant to S. B. from the Natural Science and Engineering Research Council of Canada (no. 371336).

Reprint requests should be sent to Steven Brown, Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main St. West, Hamilton, ON L8S 4K1, Canada, or via e-mail: stebro@mcmaster.ca.

### REFERENCES

- Abbott, H. P. (2008). *The Cambridge introduction to narrative* (2nd ed.). Cambridge: Cambridge University Press.
- AbdulSabur, N. Y., Xu, Y., Liu, S., Chow, H. M., Baxter, M., Carson, J., et al. (2014). Neural correlates and network connectivity underlying narrative production and comprehension: A combined fMRI and PET study. *Cortex*, *57*, 107–127.
- Andric, M., Solodkin, A., Buccino, G., Goldin-Meadow, S., Rizzolatti, G., & Small, S. L. (2013). Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia*, *51*, 1619–1629.
- Arbib, M. (2012). *How the brain got language: The mirror system hypothesis*. Oxford: Oxford University Press.
- Armstrong, D. F., & Wilcox, S. E. (2007). *The gestural origin of language*. Oxford: Oxford University Press.
- Awad, M., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. (2007). A common system for the comprehension and production of narrative speech. *Journal of Neuroscience*, *27*, 11455–11464.
- Beattie, G. (2016). *Rethinking body language: How hand movements reveal hidden thoughts*. London: Routledge.
- Biau, E., Morís Fernández, L., Holle, H., Avila, C., & Soto-Faraco, S. (2016). Hand gestures as visual prosody: BOLD responses

- to audio-visual alignment are modulated by the communicative nature of the stimuli. *Neuroimage*, *132*, 129–137.
- Bickerton, D. (1995). *Language and human behavior*. Seattle: University of Washington Press.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*, 2767–2796.
- Boyatzis, C. J., & Watson, M. W. (1993). Preschool children's symbolic representation of objects through gestures. *Child Development*, *64*, 729–735.
- Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain. *European Journal of Neuroscience*, *23*, 2791–2803.
- Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex*, *18*, 837–845.
- Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, *11*, 535–543.
- Carter, R. M., & Huettel, S. A. (2013). A nexus model of the temporal-parietal junction. *Trends in Cognitive Sciences*, *17*, 328–336.
- Catani, M., Jones, D. K., & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, *57*, 8–16.
- Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, *123*, 324–347.
- Corballis, M. C. (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.
- Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex*, *25*, 4596–4609.
- Dehghani, M., Boghrati, R., Man, K., Hoover, J., Zevin, J., Gimbel, S. I., et al. (2017). Decoding the neural representation of story meanings across languages. *Human Brain Mapping*, *38*, 6096–6106.
- Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *24*, 1742–1752.
- Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *Journal of Neuroscience*, *33*, 18906–18916.
- DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences, U.S.A.*, *109*, E505–E514.
- Elkins, J. (2001). *The domain of images*. New York: Cornell University Press.
- Fitch, W. T. (2011). The evolution of syntax: An exaptationist perspective. *Frontiers in Evolutionary Neuroscience*, *3*, 9.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*, 531–534.
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *358*, 459–473.
- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of “theory of mind”. *Trends in Cognitive Sciences*, *7*, 77–83.
- Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*, 197–208.
- Hassabis, D., Spreng, R. N., Rusu, A. A., Robbins, C. A., Mar, R. A., & Schacter, D. L. (2014). Imagine all the people: How the brain creates and uses personality models to predict behavior. *Cerebral Cortex*, *24*, 1979–1987.
- Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*, 453–458.
- Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, *3*, 272–279.
- Jackendoff, R. (2011). What is the human language faculty? Two views. *Language*, *87*, 586–624.
- Jouen, A. L., Ellmore, T. M., Madden, C. J., Pallier, C., Dominey, P. F., & Ventre-Dominey, J. (2015). Beyond the word and image: Characteristics of a common meaning system for language and vision revealed by functional and structural imaging. *Neuroimage*, *106*, 72–85.
- Kendon, A. (2015). Semiotic diversity in utterance production and the concept of ‘language’. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *369*, 20130293.
- Kestemont, J., Vandekerckhove, M., Ma, N., Van Hoeck, N., & Van Overwalle, F. (2013). Situation and person attributions under spontaneous and intentional instructions: An fMRI study. *Social Cognitive and Affective Neuroscience*, *8*, 481–493.
- Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., et al. (2009). Neural interaction of speech and gesture: Differential activations of metaphoric co-verbal gestures. *Neuropsychologia*, *47*, 169–179.
- Kreifelts, B., Ethofer, T., Shiozawa, T., Grodd, W., & Wildgruber, D. (2009). Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. *Neuropsychologia*, *47*, 3059–3066.
- Krieger-Redwood, K., Teige, C., Davey, J., Hymers, M., & Jefferies, E. (2015). Conceptual control across modalities: Graded specialisation for pictures and words in inferior frontal and posterior temporal cortex. *Neuropsychologia*, *76*, 92–107.
- Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., et al. (2010). Shared neural circuits for mentalizing about the self and others. *Journal of Cognitive Neuroscience*, *22*, 1623–1635.
- MacNeilage, P. F., & Davis, B. L. (2005). The frame/content theory of evolution of speech: A comparison with a gestural-origins alternative. *Interaction Studies*, *6*, 173–199.
- Mandler, J. M. (1984). *Stories, scripts, and scenes: Aspects of schema theory*. Hillsdale, NJ: L. Erlbaum Associates.
- Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, *62*, 103–134.
- Mason, R. A., & Just, M. A. (2006). Neuroimaging contributions to the understanding of discourse processes. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 765–799). Amsterdam: Elsevier.
- McGinn, C. (2015). *Prehension: The hand and the emergence of humanity*. Cambridge, MA: MIT Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- Nichols, S., & Stich, S. (2003). *Mindreading: An integrated account of pretense, self-awareness and understanding other minds*. Oxford: Oxford University Press.
- O'Reilly, A. W. (1995). Using representations: Comprehension and production of actions with imagined objects. *Child Development*, *66*, 309–314.
- Osaka, M., Yaoi, K., Minamoto, T., & Osaka, N. (2014). Serial changes of humor comprehension for four-frame comic Manga: An fMRI study. *Scientific Reports*, *4*, 5828.
- Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 608–620.
- Redcay, E., Velnoskey, K. R., & Rowe, M. L. (2016). Perceived communicative intent in gesture and language modulates



- the superior temporal sulcus. *Human Brain Mapping*, *37*, 3444–3461.
- Regev, M., Honey, U., & Hasson, U. (2013). Modality-selective and modality-invariant neural responses to spoken and written narratives. *Journal of Neuroscience*, *33*, 15978–15988.
- Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Brown & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 211–236). New York: Academic Press.
- Schippers, M. B., Gazzola, V., Goebel, R., & Keysers, C. (2009). Playing charades in the fMRI: Are mirror and/or mentalizing areas involved in gestural communication? *PLoS One*, *4*, e6801.
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34.
- Scott, S. K., Blank, S. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*, 2400–2406.
- Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech: A positron emission tomography study. *Journal of the Acoustical Society of America*, *120*, 1075–1083.
- Shmuelof, L., & Krakauer, J. W. (2011). Are we ready for a natural history of motor learning? *Neuron*, *72*, 469–476.
- Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences, U.S.A.*, *111*, E4687–E4696.
- Stein, N. L., & Glen, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing: Vol. 2. Advances in discourse processes* (pp. 53–120). Hillsdale, NJ: Erlbaum.
- Straube, B., Green, A., Weis, S., & Kircher, T. (2012). A supramodal neural network for speech and gesture semantics: An fMRI study. *PLoS One*, *7*, e51207.
- Suddendorf, T., Fletcher-Flinn, C., & Johnston, L. (1999). Pantomime and theory of mind. *Journal of Genetic Psychology*, *160*, 31–45.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Tallerman, M. (2015). *Understanding syntax* (4th ed.). London: Routledge.
- Tam, F., Churchill, N. W., Strother, S. C., & Graham, S. J. (2010). A new tablet for writing and drawing during functional MRI. *Human Brain Mapping*, *32*, 240–248.
- Thaler, L., & Goodale, M. A. (2011). Neural substrates of visual spatial coding and visual feedback control for hand movements in allocentric and target-directed tasks. *Frontiers in Human Neuroscience*, *5*, 92.
- Tomlin, R. S. (1986). *Basic word order: Functional principles*. London: Croom Helm.
- Van Overwalle, F., & Vandekerckhove, M. (2013). Implicit and explicit social mentalizing: Dual processes driven by a shared neural network. *Frontiers in Human Neuroscience*, *7*, 560.
- Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, *383*, 254–256.
- Vingerhoets, G., & Clauwaert, A. (2015). Functional connectivity associated with hand shape generation: Imitating novel hand postures and pantomiming tool grips challenge different nodes of a shared neural network. *Human Brain Mapping*, *36*, 3426–3440.
- Visser, M., Jefferies, E., Embleton, K. V., & Lambon Ralph, M. A. (2012). Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: Distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *Journal of Cognitive Neuroscience*, *24*, 1766–1778.
- Wilson, S., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: Intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, *18*, 230–242.
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, *42*, 654–667.
- Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences, U.S.A.*, *106*, 20664–20669.
- Yuan, Y., & Brown, S. (2014). The neural basis of mark making: A functional MRI study of drawing. *PLoS One*, *9*, e108628.
- Yuan, Y., & Brown, S. (2015). Drawing and writing: An ALE meta-analysis of sensorimotor activations. *Brain and Cognition*, *98*, 15–26.
- Zaccarella, E., & Friederici, A. D. (2017). The neurobiological nature of syntactic hierarchies. *Neuroscience & Biobehavioral Reviews*, *81*, 205–212.
- Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience*, *4*, 168.