Research Article

Psychological Science
24(8) 1379–1388

DOI: 10.1177/0956797612469209
pss.sagepub.com

Multiple Routes to Mental Animation: Language and Functional Relations Drive Motion Processing for Static Images

Kenny R. Coventry[1,2], Thomas B. Christophel[3], Thorsten Fehr[4,5,6], Berenice Valdés-Conroy[7], and Manfred Herrmann[4,5]

[1]School of Psychology, University of East Anglia; [2]Hanse Institute for Advanced Studies, Delmenhorst, Germany; [3]Bernstein Center for Computational Neuroscience, Charité-Universitätsmedizin Berlin; [4]Department of Neuropsychology and Behavioral Neurobiology, University of Bremen; [5]Center for Advanced Imaging, Bremen, Germany; [6]Department of Neurology, Otto-von-Guericke University, Magdeburg; and [7]Departamento de Psicología Básica 1, Universidad Complutense de Madrid

Abstract

When looking at static visual images, people often exhibit mental animation, anticipating visual events that have not yet happened. But what determines when mental animation occurs? Measuring mental animation using localized brain function (visual motion processing in the middle temporal and middle superior temporal areas, MT+), we demonstrated that animating static pictures of objects is dependent both on the functionally relevant spatial arrangement that objects have with one another (e.g., a bottle above a glass vs. a glass above a bottle) and on the linguistic judgment to be made about those objects (e.g., "Is the bottle above the glass?" vs. "Is the bottle bigger than the glass?"). Furthermore, we showed that mental animation is driven by functional relations and language separately in the right hemisphere of the brain but conjointly in the left hemisphere. Mental animation is not a unitary construct; the predictions humans make about the visual world are driven flexibly, with hemispheric asymmetry in the routes to MT+ activation.

Keywords

mental animation, motion processing, fMRI, language, hemispheric differences, visual perception, prediction, neuroimaging

Received 2/27/12; Revision accepted 9/18/12

One of the key features of human mental life is the ability to predict what will happen beyond the information we receive through our senses. A paradigmatic case of this ability is the inference of motion from static images. When viewing (cartoon-like) line-drawn pictures of an object in motion (Freyd, 1987), mechanical diagrams (Hegarty, 1992), or still photographs of an object in motion (Kourtzi, 2004), humans show evidence of *mental animation*—perceiving and anticipating motion beyond that portrayed by the (visual) information. Such data are consistent with the view that perception is (at least in part) about predicting what will happen, implicating a mapping between what is being perceived at any given moment in time and knowledge in memory (Bar, 2009). But what governs such mental animation?

The goal of this study was to test three possible drivers of mental animation—object knowledge, situational knowledge, and language—and possible interactions among them. We did so using functional MRI (fMRI), targeting specific regions known to be involved in motion processing. Perceived motion in humans is associated with a cluster of brain regions at the temporo-parietal-occipital junction, particularly the middle temporal and middle superior temporal areas (MT+; Dupont, Orban, De Bruyn, Verbruggen, & Mortelmans, 1994; Tootell et al.,

Corresponding Author:
Kenny R. Coventry, School of Psychology, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, England
E-mail: k.coventry@uea.ac.uk

1995). Studies using fMRI have found that viewing static images with implied motion (e.g., a picture of an athlete in an action pose) is associated with increased MT+ activation compared with viewing static images with no such implied motion (e.g., a picture of an athlete standing still; Kourtzi & Kanwisher, 2000; Senior et al., 2000). Furthermore, it has been shown that such mental animation can be eliminated when MT+ function is disrupted using transcranial magnetic stimulation (Senior, Ward, & David, 2002).

The first possible driver of mental animation that we considered was object knowledge. In the cognitive sciences, objects are accorded a privileged status, whether these objects are perceptual (the visual percept of a bottle), conceptual (the concept of a bottle), or linguistic (the noun *bottle*). These objects reflect the regularity with which specific properties in the world co-occur. For example, [solid object + made of glass + container + liquid + spout + pouring] are taken to form a stable set of features that together constitute the object representation (visual: Kahneman & Treisman, 1992; Ullman, 1996; conceptual: Margolis & Laurence, 1999; or lexical: Pustejovsky, 1995) for "bottle." Consistent with this object-knowledge account, it has been shown that labeling the same object in different ways (e.g., an ambiguous geometric shape labeled "rocket" vs. "steeple") affects memory for the last position of that object when it is seen moving along an expected path (Reed & Vinson, 1996).

A second possible driver of mental animation is knowledge of how objects interact in situations. Visual objects co-occur with other visual objects, and they do so with varying likelihoods. Both adults and infants are sensitive to these regularities and associated likelihoods across objects in learning about the visual world (Turk-Browne, Scholl, Chun, & Johnson, 2009). And it has been shown that an object is more easily detected in a visual scene when it co-occurs with familiar objects in spatially congruent locations than when it co-occurs with the same objects in spatially incongruent locations (Biederman, 1972; Henderson & Hollingworth, 1999). Moreover, in some cases, individual features in the world might be linked more profitably to situations in which multiple objects occur rather than to situations with individual objects. For example, pouring, often taken as a feature of a bottle, usually occurs with a receptacle present, and the likelihood of pouring is presumably greater when a receptacle is present than when it is absent.

A third—and particularly intriguing—source of information in memory that may affect mental animation is co-occurrence relations across domains, in particular, the binding together of language and perception. Just as people are sensitive to how visual features cluster together and to how objects co-occur, so-called grounded, or embodied, views of cognition (Barsalou, 2008; Gallese & Lakoff, 2005; Rizzolatti & Arbib, 1998) assume tight coupling between language and nonlinguistic systems. In support of such theories, it has been shown that perceptual and motor-system activations (*perceptual simulations*) occur during language processing. For example, early motor activations have been found when reading words and sentences involving action verbs (reading *kick* activates motor areas involved in performing kicking actions; Boulenger et al., 2006; Grèzes & Decety, 2001; Pulvermüller, Shtyrov, & Ilmoniemi, 2005). However, the mechanisms underlying the grounding of language in nonlinguistic systems are not well understood. One possibility is that the likelihood with which words co-occur with perceptual events during learning may give rise to differential degrees of perceptual activations in later language processing (consistent with a Hebbian approach to learning; Pulvermüller, 2001). By the same token, such an account also predicts that differential activations in visual processing may occur as a function of the language presented with those visual images.

To test these possible drivers of mental animation and the interactions among them, we had participants complete a sentence-picture verification task in an fMRI scanner, using a 3 (picture condition) × 3 (language condition) design. We manipulated the pictures to examine whether individual objects or knowledge of how objects interact in context drives motion processing of static images. Consider an image of a bottle positioned higher than a glass versus an image of a bottle positioned lower than a glass. If learning about these objects involves knowing that they interact, then simulating the path of liquid falling from the bottle should occur only when the glass is positioned below the bottle. Thus, MT+ activation should be greater when the two objects are in functionally congruent positions, such that the bottle can interact with the glass, than when the same objects are present but in positions in which pouring would not be a relevant interaction. In contrast, if knowledge of objects drives motion processing, then there should be no difference in MT+ activation whether the bottle is positioned higher or lower than the glass.

To establish whether language drives or mediates MT+ activation, we manipulated the relational term used in sentences for the same objects. The comprehension of sentences involving spatial prepositions (e.g., *in, over, near*) is affected by knowledge about how objects are interacting or will interact with each other as well as knowledge about where those objects are positioned in relation to each other (Coventry & Garrod, 2004). *The bottle is over the glass* is judged as being a more appropriate description of a picture of a bottle and glass when water protruding from the spout of the bottle is shown (or expected) to end up in the glass rather than to miss the glass (Coventry et al., 2010; Coventry & Garrod,

2004). Hence, for static images involving containers beginning to pour liquids or objects toward other containers, we predicted that spatial-language judgments (e.g., "Is the bottle over the glass?") would require motion processing of those static images to establish whether one container is indeed over the other. In contrast, judgments involving comparative adjectives (e.g., "Is the bottle bigger than the glass?") require processing the size of objects, and, accordingly, motion processing may not be necessary for such judgments of the same pictures.

In addition, we were interested in establishing whether language mediates premotor and motor activations associated with picture processing. Understanding whether a bottle is over a glass may require not only animating the path of liquid from the bottle to the glass but also simulating the action of pouring. It has been shown that viewing objects automatically potentiates actions toward those objects (see, e.g., Grèzes, Tucker, Armony, Ellis, & Passingham, 2003; Tucker & Ellis, 1998). We wondered whether such motor activations are dependent on functional relations between objects (consistent with data from patients with visual extinction; Humphreys & Riddoch, 2007) and, furthermore, whether the relevance of the action cued by language affects the extent of such activation patterns.

Method

Participants

Twelve healthy, right-handed native English speakers (9 males, 3 females; mean age = 30.25 years, range = 21–45 years) recruited from the city of Bremen, Germany, participated in the experiment. All participants grew up in an English-speaking country with English as their first language and had moved to Germany only recently. Prior to the recruitment of participants, the study was given a full ethical scrutiny and was subsequently approved in accordance with international ethical standards.

Experimental procedure and materials

Participants completed a sentence-picture verification task while in an fMRI scanner. The task was to judge whether presented sentences with the form *The [Noun 1] is [term] the [Noun 2]* were true or false descriptions of subsequently presented pictures (see Fig. 1a). Across three language conditions, sentences contained one of two types of closed-class terms: a spatial preposition (vertical prepositions: *over, under, above, below*; proximity prepositions: *near, far*) or a comparative adjective (*bigger, smaller*). Pictures were of three types (see Figs. 1b, 1c, and 1d; see also Fig. S1 in the Supplemental Material available online). In two of the picture conditions, two objects (Noun 1 and Noun 2) that frequently appear together were shown either in functionally congruent (FC) positions (e.g., a packet positioned higher than a pan; see Fig. 1b) or in functionally incongruent (FI) positions (e.g., a pan positioned higher than a packet; see Fig. 1c). In a third condition, nonfunctional (NF) objects that do not usually co-occur and do not interact functionally were presented (e.g., a pepper positioned higher than a telephone; see Fig. 1d). Both the FC and FI pictures showed a container with falling objects protruding from its spout or edge at various angles (see Fig. S1). In all conditions, objects were matched for size, distance, and position.

There were 216 trials in the sentence-picture verification task—24 trials for each of the nine conditions (3 picture conditions × 3 language conditions). These trials were matched for object position (Fig. S1d), mirroring (Fig. S1e), and, in the FC and FI conditions, the position and angle of falling objects. We manipulated the noun order and preposition direction in sentences (e.g., "The bottle is over the glass" vs. "The glass is under the bottle" vs. "The glass is over the bottle"), ensuring that on 8 of every 24 trials, the descriptions were false (so that participants would attend throughout the task).

After participants had provided written consent, they were instructed that their task was to indicate, by pressing a button as quickly as possible, whether a picture was correctly described by the sentence that preceded it. The intertrial interval was jittered between 0.65 s and 1.36 s. A practice run comprised eight (random) trials from the main experiment. Data for baseline and sentence-picture verification trials were acquired in a single continuous scanning run (see Fig. S2).

The order of the conditions in the sentence-picture verification task was determined by a pseudorandomized, nonstationary probabilistic design (Friston, 2000; see Fig. S2). Performance on a simple attentional task using stimuli comprising geometric objects matched in size and color was recorded as a baseline (see Fig. S1c); in this task, participants were instructed to press any button when an object appeared on the screen. In addition, we used a localizer scan to individually define functional regions of interest (ROIs) for MT+. The localizer paradigm consisted of squares either floating into the screen (radial-flow field with size change) or remaining static (static field), consistent with localizers used previously (Tootell et al., 1995).

fMRI recording and analysis

Functional blood oxygen level–dependent data were acquired with a 3-T Magnetom Allegra head scanner that uses a whole-head, local-gradient coil (Siemens, Erlangen, Germany). Functional images were collected with an echoplanar imaging sequence during the sentence-picture

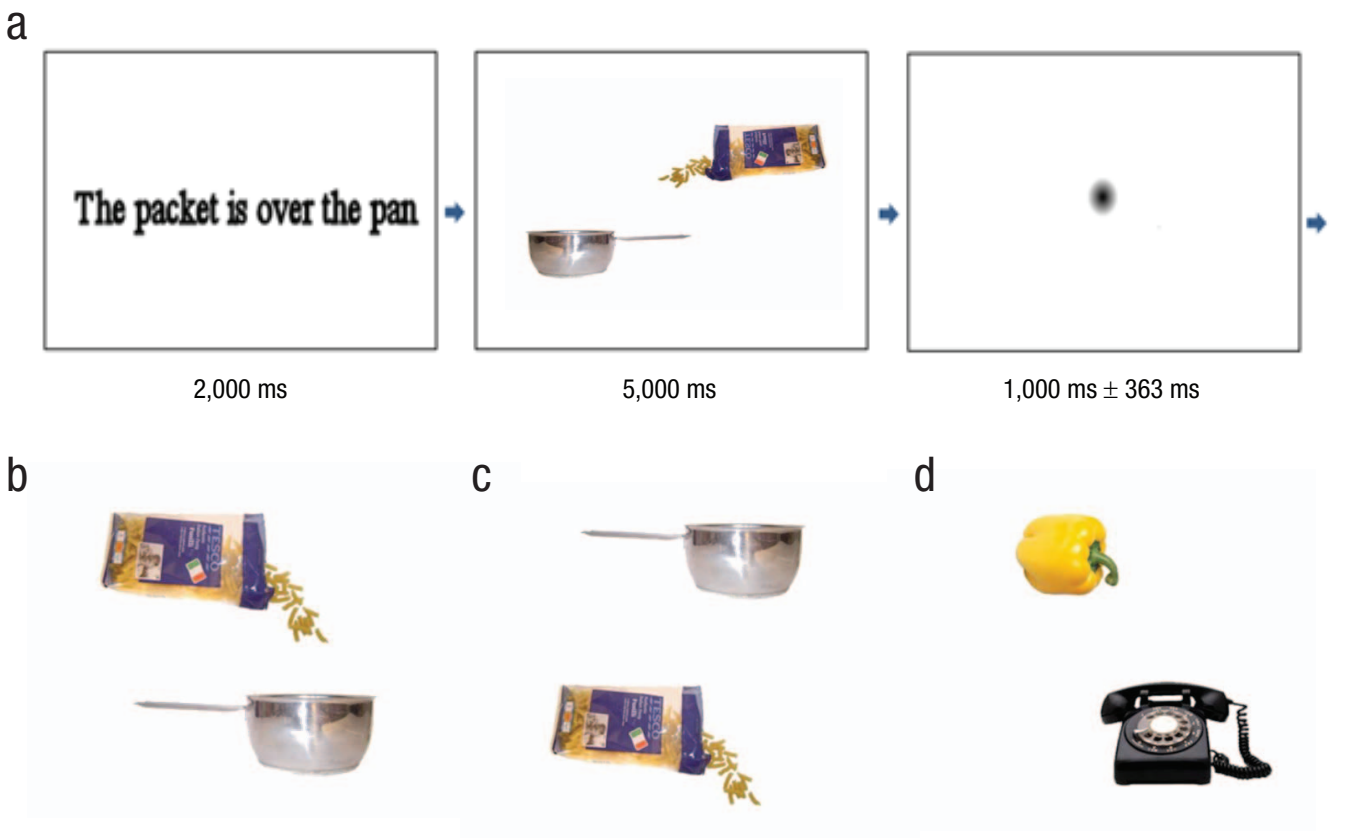

Fig. 1. Example trial sequence and stimuli from the sentence-picture verification task. On each trial (a), sentences containing prepositions (vertical prepositions, i.e., *over, under, above,* or *below*—or proximity prepositions—*near* or *far*) or comparative adjectives (*bigger* or *smaller*) describing a relation between two objects were presented to participants. A picture of two objects was then presented. Participants had to indicate whether the sentence accurately described the picture. Pictures were of three types: *Functionally congruent* pictures (b) showed two objects that frequently appear together in functionally congruent positions; *functionally incongruent* pictures (c) showed two objects that frequently appear together in functionally incongruent positions; and *nonfunctional* pictures (d) showed two objects that do not usually co-occur and do not interact functionally. The nonfunctional objects were matched for size, shape, and color to the objects in the other conditions. For all three picture types, the distances between objects were manipulated, and for the functional pictures, the expected trajectory of the falling objects was also varied (see Fig. S1e in the Supplemental Material available online).

verification task, the baseline task, and the localizer task (repetition time = 2.06 s, echo time = 30 ms). A total of 38 slices of 3-mm^3 voxels were acquired across the whole cerebral cortex. High-resolution T1-weighted structural scans were acquired as a set of 160 contiguous saggital slices (1 × 1 × 1 mm voxels; magnetization-prepared rapid-gradient-echo repetition time = 2.3 s, echo time = 4.38 ms; 256 × 256 matrix, flip angle 8°).

SPM5, MarsBaR, and STATISTICA were used for analysis. After correcting for slice-acquisition order and head motion, the functional data were coregistered with the structural scan. Normalizing parameters were obtained using the unified-segmentation approach, and we normalized the functional data onto Montreal Neurological Institute (MNI) space. Finally, the normalized data were smoothed with a Gaussian kernel of 9 mm (full width at half maximum). The left and right MT+ were localized for each participant as the regions responding more strongly to flow-field than to static-field stimulation (Culham, He, Dukelow, & Verstraten, 2001). We calculated percentage signal change for all experimental conditions as computed by MarsBaR for MT+ in both hemispheres. The percentage-signal-change data were incorporated into a Sentence Type × Picture Type repeated measures analysis of variance (ANOVA).

We also investigated whether areas other than MT+ showed consistent changes in activation in our 3 (sentence type) × 3 (picture type) design using a mass-univariate approach. We created contrast images for the nine conditions (one per cell of the design) against baseline and incorporated these images into a second-level between-subjects 3 × 3 ANOVA (N = 12). We then tested for main effects of picture and language and for a Picture × Language interaction, all ps < .05 (false discovery rate,

FDR, controlled), $k = 20$, and compared single conditions for the cluster peaks. We successively created an overlay with the activation maps for the main effects of language and picture areas at the temporo-parietal-occipital junction that showed significant activations in the MT+ localizer, $p < .05$ (uncorrected).

Results

fMRI analyses

To test for possible effects of picture condition and language condition, we performed both ROI analyses (for MT+) and whole-brain analyses to test for other potential differences in brain-region activation.

After successfully localizing MT+ in 10 participants (see Fig. 2a), we compared mean percentage signal change (over baseline) of MT+ activation for left and right MT+ separately using 3 (sentence condition) × 3 (picture condition) ANOVAs.

For left MT+ and right MT+, there were significant effects of picture condition—left hemisphere: $F(2, 18) = 13.7$, $p < .0001$, $\eta_p^2 = .603$; right hemisphere: $F(2, 18) = 3.70$, $p < .05$, $\eta_p^2 = .291$ (Fig. 2b). In both hemispheres, the FC picture condition was associated with significantly greater MT+ activation compared with either of the other conditions (all ps < .05), whereas MT+ activation did not differ significantly between the FI and NF conditions (both ps > .4). For left MT+, there was no main effect of language condition, but there was an interaction between sentence condition and picture condition, $F(4, 36) = 2.56$, $p = .05$, $\eta_p^2 = .222$ (Fig. 2c). In both of the spatial-preposition conditions, the FC pictures showed significantly greater activation than the FI and NF pictures (all ps < .01), whereas there was no difference between the FI and NF pictures. No differences were present across any of the picture conditions in which comparative adjective judgments were made (all ps > .05). For right MT+, this interaction was not present, $F < 1$ (Fig. 2c), but there was a main effect of sentence condition, $F(2, 18) = 5.63$, $p < .05$, $\eta_p^2 = .385$, such that there was greater activation overall for the spatial-preposition conditions than for the comparative-adjective condition. These findings were confirmed in a further MT+ Hemisphere × Picture × Language ANOVA, which revealed a reliable three-way interaction, $F(2, 18) = 3.63$, $p < .05$, $\eta_p^2 = .287$.

We also investigated whether areas other than MT+ showed consistent changes in activation in our 3 (sentence type) × 3 (picture type) design using a voxel-wise mass-univariate ANOVA (whole-brain analysis). There were main effects of the language manipulation in multiple areas, all ps < .05 (FDR controlled), $k = 20$ (see Table 1). Consistent with imaging work on spatial-relation processing and spatial-language processing (Amorapanth, Widick, & Chatterjee, 2010; Damasio et al., 2001; Noordzij, Neggers, Ramsey, & Postma, 2008; Wallentin, Ostergaard, Lund, Ostergaard, & Roepstorff, 2005), there was increased activation in the posterior parietal cortex on trials with sentences containing spatial prepositions as compared with trials with sentences containing comparative adjectives: right posterior parietal cortex (MNI peak coordinates: $x = 18$, $y = 56$, $z = 20$), $t(11) = 6.74$, $p < .001$; left posterior parietal cortex (MNI peak coordinates: $x = 6$, $y = 58$, $z = 56$), $t = 5.08$, $p < .001$. In contrast, trials with comparative adjectives induced increased activations in early and ventral visual areas: right occipital lobe (MNI peak coordinates: $x = 26$, $y = 94$, $z = 6$), $t(11) = 6.51$, $p < .001$; left early and ventral visual areas (MNI peak coordinates: $x = 36$, $y = 88$, $z = 10$), $t(11) = 5.93$, $p < .001$.

We also found a main effect of language bilaterally at the temporo-parietal-occipital junction (Fig. 3), which is usually seen as the location of MT+ (Culham et al., 2001). This activation difference is attributable to increased activation in the spatial-preposition conditions relative to the comparative-adjective condition: right temporo-parietal-occipital junction (MNI peak coordinates: $x = 58$, $y = 56$, $z = 4$), $t(11) = 5.17$, $p < .001$; left temporo-parietal-occipital junction (MNI peak coordinates: $x = 54$, $y = 52$, $z = 14$), $t(11) = 4.54$, $p < .001$. In the same comparison, activity was increased in the right supplementary motor area (MNI peak coordinates: $x = 14$, $y = -4$, $z = 64$), $t(11) = 3.5$, $p < .001$, and in the left premotor cortex (MNI peak coordinates: $x = -48$, $y = 4$, $z = 50$), $t(11) = 2.4$, $p < .05$. No differences in brain activation were present between the two spatial-preposition conditions. For the main effect of picture, we found only one significant cluster at the left temporo-parietal-occipital junction (MT+; see Fig. 3). Activity in this area was higher in the FC condition than in both the FI and NF conditions (MNI peak coordinates: $x = -54$, $y = -74$, $z = 0$), $t(11) = 4.89$, $p < .001$. There was no difference between the FI and NF conditions. No voxels showed a significant interaction effect after correction for multiple comparisons, $p < .05$ (FDR controlled).

Because MT+ activity can be modulated by eye movements and visual attention (Dukelow et al., 2001; O'Craven, Rosen, Kwong, Treisman, & Savoy, 1997), it is important to note that there was no main effect of picture with respect to the frontal eye fields. Saccadic activity generally is accompanied by increased activation in the frontal eye fields (Luna et al., 1998); thus, the picture contrasts (or the Language × Picture × Hemisphere interaction) cannot be explained by differences in eye-movement patterns.

Behavioral data

We calculated the number of "yes" responses (i.e., when the sentence preceding the picture was thought to be a correct description of the picture) for each participant for

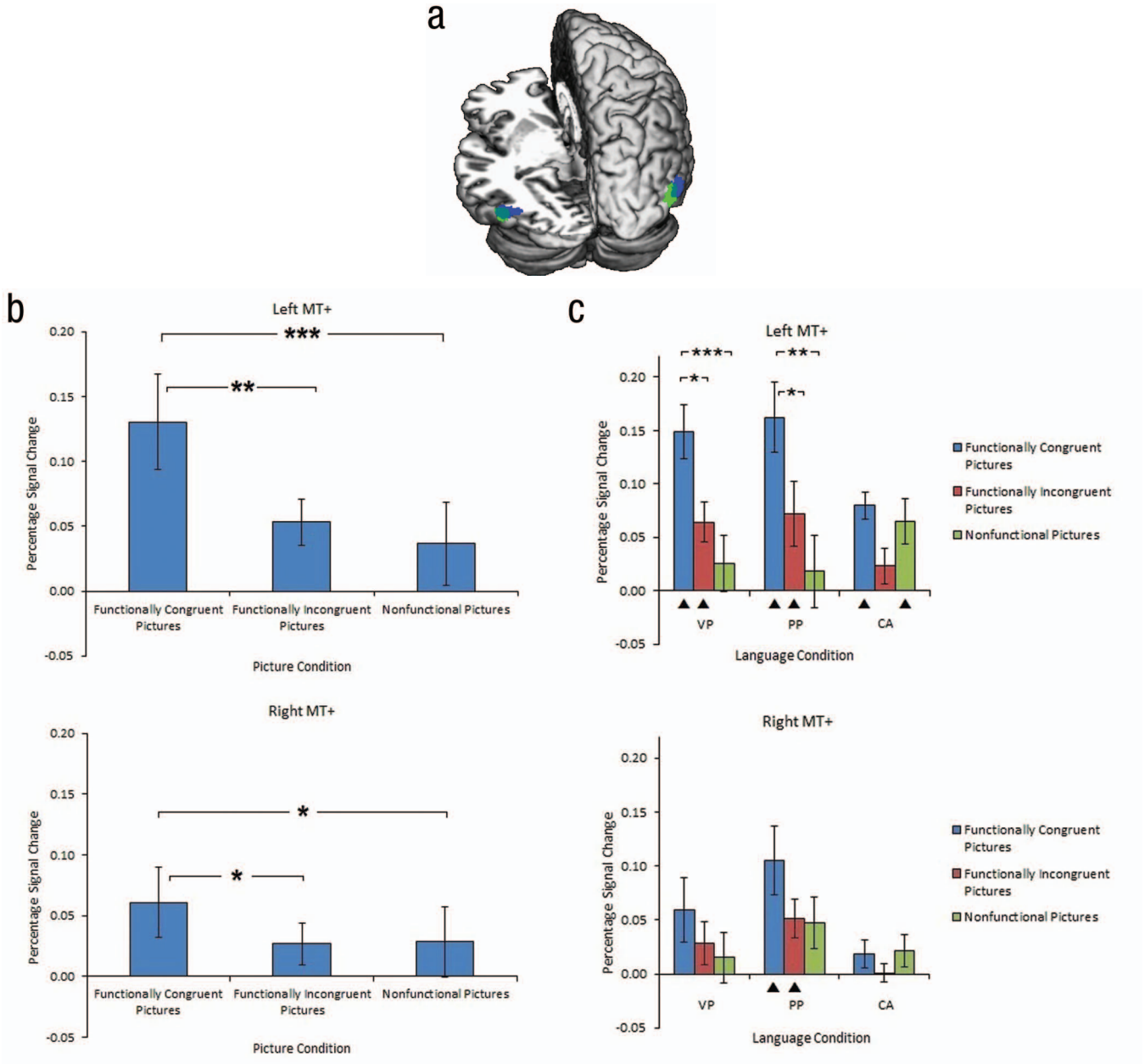

Fig. 2. Activations and percentage signal changes in the left and right middle temporal and middle superior temporal areas (MT+). The rendered brain image (a) shows localized MT+ activations for two representative participants (green and blue, respectively). The graphs in (b) show mean percentage signal change as a function of picture condition (results of region-of-interest, ROI, analyses). The graphs in (c) show mean percentage signal change as a function of language condition and picture condition (results of ROI analyses). In (b) and (c), error bars show standard errors of the mean. Triangles in (c) show conditions with percentage-signal-change values significantly greater than zero (one-sample *t* tests). VP = vertical-preposition (*over, under, above, below*) condition; PP = proximity-preposition (*near, far*) condition; CA = comparative-adjective (*bigger, smaller*) condition. Asterisks indicate significant differences between conditions (*$p < .05$, **$p < .01$, ***$p < .001$).

the FC and FI images preceded by sentences with spatial prepositions. In particular, we wanted to test whether the position of falling objects (see Fig. S1e) affected responses. In a previous study (Coventry et al., 2010), participants drew the expected continuation of the falling objects for all four positions and two angles (the functional and nonfunctional cases) in the FC scenes (see Figs. S1d and S1e). On the basis of these responses from the previous

Table 1. Active Clusters for the Main Effect of Language

Region and area	Hemisphere	Brodmann's area	MNI coordinates			Main effect		SP > CA	
			x	*y*	*z*	$F(2, 22)$	$p_{(FDR)}$	$t(11)$	$p_{(uncorrected)}$
Temporo-parietal-occipital junction	Right	22	58	−56	4	14.01	.001	5.17	< .001
	Left	22	−54	−52	14	12.4	.002	4.54	< .001
Motor									
Supplementary motor area	Right	6	14	−4	64	7.65	.021	3.5	< .001
Premotor	Left	6	−48	4	50	7.44	.024	2.4	.016
Parietal									
Precuneus	Right	7/31	18	−56	20	24.55	< .001	6.74	< .001
	Left	7/31	−6	−58	56	17.44	< .001	5.08	< .001
Angular gyrus	Right	39	48	−74	32	22.68	< .001	6.49	< .001
Inferior lobule	Right	40	70	−32	30	11.6	.002	3.68	< .001
	Left	40	−64	−42	34	8.59	.012	3.73	< .001
Occipital									
Middle gyrus	Right	18	26	−94	6	21.74	< .001	−6.51	< .001
	Left	18	−36	−88	−10	17.81	< .001	−5.93	< .001
Superior gyrus	Left	19	−36	−84	32	10.45	.004	4.38	< .001
Prefrontal									
Inferior gyrus	Right	46	52	40	14	10.27	.005	−2.69	.008
Medial gyrus	Right	10	6	54	−6	8.78	.011	4.17	< .001
	Left	10	−2	58	−4	10.2	.005	4.35	< .001
Middle gyrus	Left	9	−24	26	34	8.24	.015	4.04	< .001
Temporal									
Inferior gyrus	Left	21	−60	−6	−20	7.97	.017	3.99	< .001
Cingulate gyrus	Left	24	−2	−16	34	12.56	.001	4.72	< .001

Note: Anatomical labels and Brodmann's areas were assigned using anatomical-labeling software (Lancaster et al., 2000) and visual inspection of the full clusters. Coordinates of the peaks are noted in stereotaxic Montreal Neurological Institute (MNI) space. F and $p_{(FDR)}$ (false-discovery rate controlled) values are based on the main effect of language; T and $p_{(uncorrected)}$ values are based on a contrast of the spatial-preposition (SP) and comparative-adjective (CA) conditions. Negative t values indicate that a cluster responded most strongly during trials that included comparative adjectives.

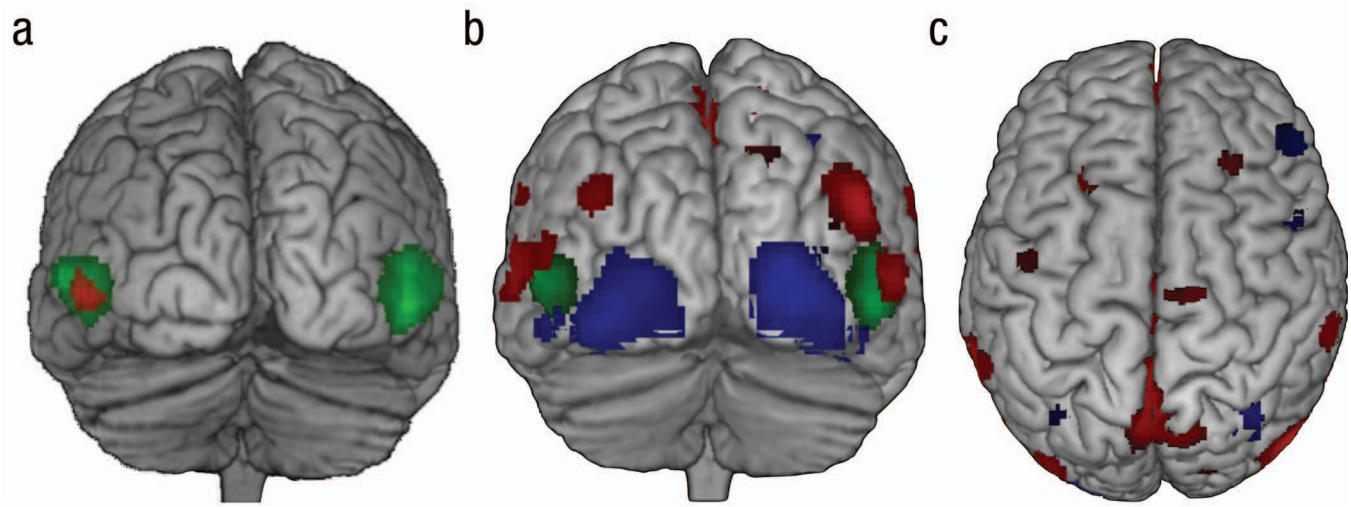

Fig. 3. Activation differences between the picture and language conditions (results of whole-brain analyses). The three renderings of human brains illustrate (a) differences in activation showing the main effect of picture condition, $p_{(FDR)}$ (false-discovery rate controlled) < .05, k = 20, and (b and c) differences in activation showing a main effect of language condition, $p_{(FDR)}$ < .05, k = 20. Voxels with higher activation in the spatial-preposition conditions than in the comparative-adjective condition are shown in red, whereas the opposite contrast is shown in blue. Panels (a) and (b) include a functional between-subjects region of interest for middle temporal and middle superior temporal areas (MT+; shown in green).

study, we established whether participants agreed with our prior classification of scenes as functional or nonfunctional. For functional scenes, participants' drawn trajectories should show the falling objects reaching the bottom object, whereas trajectories should miss the bottom object for nonfunctional scenes. Results showed high agreement with our prior classifications for vertically displaced scenes (93%) but showed low agreement for horizontally displaced scenes (23%). All horizontally displaced items were therefore removed from the behavioral analyses. (It was not necessary to remove these items in the fMRI analyses because all the scenes required mental animation to establish whether or not the falling objects would end up in the container below.)

A 2 (spatial-preposition type: vertical or proximity) × 2 (picture condition: FC or FI) × 2 (position of falling objects: functional or nonfunctional) × 2 (vertical position: near or far) within-subjects ANOVA produced a number of main effects and interactions indicating that participants took the path of falling objects into account when making their judgments. In particular, there was an interaction between picture condition and the position of falling objects, $F(1, 10) = 9.073$, $p = .013$. For the FC pictures, scenes in which the falling objects were expected to end up in the container elicited more "yes" responses ($M = 0.818$) than did those in which the falling objects were expected to miss the container ($M = 0.739$), $p = .04$. This was not the case for the FI picture condition, in which the nonfunctional scenes ($M = 0.875$) elicited more "yes" responses than the functional scenes did ($M = 0.784$), $p = .04$.

Equivalent analyses of the comparative-adjective judgments revealed no significant main effects or interactions for these terms.

We also analyzed reaction times in a 3 (language condition: vertical preposition, proximity preposition, or comparative adjective) × 2 (picture condition: FC or FI) × 2 (position of falling objects: functional or nonfunctional) × 2 (vertical position: near or far) within-subjects ANOVA. There was an interaction between language condition and position of falling objects, $F(2, 20) = 6.992$, $p = .005$. Reaction times for near positions were faster than for far positions for the proximity-preposition condition ($p < .001$, Bonferroni contrast), but this was not the case for either of the other two language conditions (both $ps > .05$).

In summary, the behavioral data revealed that participants considered the potential path of falling objects when viewing pictures preceded by sentences containing spatial prepositions but not when viewing the same pictures preceded by sentences containing comparative adjectives. Response speed was not influenced by picture condition or by the position of falling objects.

General Discussion

The results show that mental animation for pouring events is driven not by individual objects but by expectations regarding how objects typically interact. This is consonant not only with views of cognition as situated action (Barsalou, 2008) but also with the view that perceptual objects are determined by the frequency with which features in the world co-occur (Humphreys & Riddoch, 2007). Pouring more frequently occurs with a bottle and glass in a particular spatial relation than with a bottle alone (e.g., bottles alone are often seen sitting on shelves), and this type of knowledge drives mental animation.

Language also plays an important role in determining the extent of mental animation. In right MT+, there was a main effect of language, with increased activation for the spatial-preposition conditions compared with the comparative-adjective condition. In left MT+, the effect of language was not reliable; the interaction showed an effect of picture type when pictures were preceded by sentences containing spatial prepositions but not when pictures were preceded by sentences containing comparative adjectives. This pattern was not found in right MT+. We take these results as strong support for the view that the way in which language co-occurs with objects affects the types of mental simulations performed when viewing those objects. Furthermore, the interaction on the left side alone suggests that the binding together of language and visual events is primarily computed in the left hemisphere, consistent with recent work showing lateralized effects of language on categorical perception (Gilbert, Regier, Kay, & Ivry, 2006).

Results of the MT+ ROI analyses were generally supported in the whole-brain analyses, with main effects of language condition and picture condition. The overall coarser results (e.g., the absence of the interaction in left MT+) are to be expected, given the reduced power of whole-brain analyses due to anatomical rather than functional region identification and the rather conservative criteria adopted for significance (see Saxe, Brett, & Kanwisher, 2006, for a discussion of limitations in anatomical averaging).

The whole-brain analyses also demonstrated that language mediates motor activations when people view pictures that follow sentences, such that motor activations after viewing sentences containing spatial prepositions are greater than motor activations after viewing sentences containing comparative adjectives. In contrast, we did not find any picture-condition differences with respect to motor activations, which suggests that such effects may be tied to objects in the case of motor affordances rather than object interactions (e.g., seeing an object that one

can grasp affords grasping). These motor activations are consistent with motor theories of action and language processing (Gallese & Lakoff, 2005; Pulvermüller, 2001; Rizzolatti & Arbib, 1998), but our results indicate that presenting different terms with the same objects is enough to switch motor simulation on and off.

Our results support a Hebbian approach to learning and later activations. The likelihood with which particular combinations of relations co-occur during learning directly affects what activations occur during retrieval. Moreover, the different results for left and right MT+ suggest that there are multiple routes driving mental animation: The right MT+ represents the likelihood with which perceptual objects in particular configurations and particular types of language separately co-occur with motion events, and the left MT+ represents the likelihood with which particular types of language and perceptual-object combinations conjointly occur with dynamic events. Mental animation is not a unitary construct; the predictions humans make about the visual world benefit from flexible routes to meaning construction.

Author Contributions

K. R. Coventry conceived the study and took primary responsibility for the preparation of this manuscript. K. R. Coventry and T. Christophel designed the study with input from T. Fehr, M. Herrmann, and B. Valdés-Conroy. B. Valdés-Conroy prepared the visual stimuli. T. Christophel ran the experiment and took primary responsibility for neuroimaging analyses and the drafting of neuroimaging results, with advice and assistance from T. Fehr, M. Herrmann, and K. R. Coventry. K. R. Coventry analyzed the behavioral data. All authors commented on drafts of the manuscript.

Acknowledgments

The authors thank Paul Engelhardt, Andreas Finkelmeyer, Debra Griffiths, Franz Mechsner, and Larry Taylor for commenting on an earlier draft of this manuscript and Evelyn Ferstl for helpful discussions in the early stages of the project.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by a Hanse-Wisschenschaftskolleg fellowship (awarded to K. R. C.) and by Center for Advanced Imaging Bremen Grant BMBF01GO0506 (awarded to M. H.).

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

References

Amorapanth, P. X., Widick, P., & Chatterjee, A. (2010). The neural basis for spatial relations. *Journal of Cognitive Neuroscience*, *22*, 1739–1753.

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, *364*, 1235–1243.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*, 77–80.

Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., & Nazir, T. A. (2006). Cross-talk between language processes and overt motor behavior in the first 200 ms of processing. *Journal of Cognitive Neuroscience*, *18*, 1607–1615.

Coventry, K. R., & Garrod, S. C. (2004). *Saying, seeing and acting: The psychological semantics of spatial prepositions*. Hove, England: Psychology Press.

Coventry, K. R., Lynott, D., Cangelosi, A., Monrouxe, L., Joyce, D., & Richardson, D. C. (2010). Spatial language, visual attention, and perceptual simulation. *Brain and Language*, *112*, 202–213.

Culham, J., He, S., Dukelow, S., & Verstraten, F. A. J. (2001). Visual motion and the human brain: What has neuroimaging told us? *Acta Psychologica*, *107*, 69–94.

Damasio, H., Grabowski, T. J., Tranel, T., Ponto, L. L. B., Hichwa, R. D., & Damasio, A. R. (2001). Neural correlates of naming actions and of naming spatial relations. *NeuroImage*, *13*, 1053–1064.

Dukelow, S. P., DeSouza, J. F. X., Culham, J. C., van den Berg, A. V., Menon, R. S., & Vilis, T. (2001). Distinguishing subregions of the human MT+ complex using visual fields and pursuit eye movements. *Journal of Neurophysiology*, *86*, 1991–2000.

Dupont, P., Orban, G. A., De Bruyn, B. A., Verbruggen, A., & Mortelmans, L. (1994). Many areas in the human brain respond to visual motion. *Journal of Neurophysiology*, *72*, 1420–1424.

Freyd, J. (1987). Dynamic mental representations. *Psychological Review*, *94*, 427–438.

Friston, K. J. (2000). Experimental design and statistical issues. In J. C. Mazziotta & A. W. Toga (Eds.), *Brain mapping: The disorders* (pp. 33–58). San Diego, CA: Academic Press.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*, 455–479.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences, USA*, *103*, 489–494.

Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, *12*, 1–19.

Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate actions: An

fMRI study of implicit processing. *European Journal of Neuroscience*, *17*, 2735–2740.

Hegarty, M. (1992). Mental animation: Inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1084–1102.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.

Humphreys, G. W., & Riddoch, M. J. (2007). How to define an object: Evidence from the effects of action on perception and attention. *Mind & Language*, *22*, 534–547.

Kahneman, D., & Treisman, D. A. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219.

Kourtzi, Z. (2004). But still, it moves. *Trends in Cognitive Sciences*, *8*, 47–49.

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

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., . . . Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, *10*, 120–131.

Luna, B., Thulborn, K. R., Strojwas, M. H., McCurtain, B. J., Berman, R. A., Genovese, C. R., & Sweeney, J. A. (1998). Dorsal cortical regions subserving visually guides saccades in humans: An fMRI study. *Cerebral Cortex*, *8*, 40–47.

Margolis, E., & Laurence, S. (Eds.). (1999). *Concepts*. Cambridge, MA: MIT Press.

Noordzij, M. L., Neggers, R. H. J. B., Ramsey, N., & Postma, A. (2008). Neural correlates of locative prepositions. *Neuropsychologia*, *46*, 1576–1580.

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*, 591–598.

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, *5*, 517–524.

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, *17*, 1–9.

Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.

Reed, C. L., & Vinson, N. G. (1996). Conceptual effects on representational momentum. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 839–850.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neurosciences*, *21*, 188–194.

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *NeuroImage*, *30*, 1088–1096.

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

Senior, C., Ward, J., & David, A. S. (2002). Representational momentum and the brain: An investigation into the functional necessity of V5/MT. *Visual Cognition*, *9*, 81–92.

Tootell, R. B. H., Reppas, J. B., Dale, A. M., Look, R. B., Sereno, M. I., Malach, R., . . . Rosen, B. R. (1995). Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature*, *375*, 139–141.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential action. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 830–846.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*, 1934–1945.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA: MIT Press.

Wallentin, M., Ostergaard, S., Lund, T. E., Ostergaard, L., & Roepstorff, A. (2005). Concrete spatial language: See what I mean? *Brain and Language*, *92*, 221–233.