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

Kenny R. Coventry[1, 2, 3], Thomas Christophel[4], Thorsten Fehr[5, 6, 7], Berenice Valdés-Conroy[8], & Manfred Herrmann[5, 6]

[1]School of Psychology, University of East Anglia, UK

[2]Cognition and Communication Research Centre, Northumbria University, UK

[3]Hanse Institute for Advanced Studies, Germany

[4]Bernstein Center for Computational Neuroscience, Charité-Universitätsmedizin Berlin, Germany

[5] Department of Neuropsychology and Behavioral Neurobiology, University of Bremen, Germany

[6]Center for Advanced Imaging (CAI) – Bremen, Germany

[7]Department of Neurology II, Otto-von-Guericke University, Magdeburg, Germany

[8] Departamento de Psicología Básica I, Universidad Complutense de Madrid, Spain

Address for Correspondence:

Kenny R. Coventry, School of Psychology, Faculty of Social Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom. E-mail: k.coventry@uea.ac.uk

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; area MT+), we demonstrate that animating static pictures of objects is dependent both on the functionally-relevant spatial arrangement objects have with one another (e.g. *a bottle above a glass* versus *a glass above a bottle*), and the linguistic judgment to be made about those objects (e.g. *Is the bottle above the glass?* versus *Is the bottle bigger than the glass?*). Furthermore, we show that mental animation is driven by functional relations and language separately in the right hemisphere, but in interaction in the left brain. 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.

Key words: mental animation; motion processing; fMRI; language; hemispheric differences

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 is inferring 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 ahead of the (visual) information given. 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 between them. We did so using functional magnetic resonance imaging (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-parieto-occipital junction; particularly the middle temporal and middle superior temporal areas – hereafter MT+(Dupont et al., 1994; Tootell et al., 1995). Using fMRI, it has also been found that viewing static images with implied motion (e.g., a picture of an athlete in a moving pose) is associated with increased MT+ activation compared to pictures with no such implied motion (e.g., 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 et al., 2002).

The first possible driver of mental animation we considered was object knowledge. In the cognitive sciences, objects are accorded a privileged status, whether those objects are perceptual (visual percept a bottle, etc.), 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 comprise a stable set of features that together constitute the object representation (visual (Kahneman & Treisman, 1992; Ullman, 1996), conceptual (Margolis & Laurence, 1999), lexical (Pustejovsky, 1995)) BOTTLE. Consistent with this object knowledge account, it has been shown that labeling the same object in different ways (e.g. a 'rocket' versus a 'cathedral') affects memory for the last position of that object when it is seen moving along its 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 et al., 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 their spatially congruent locations than when it co-occurs with those same objects in spatially incongruent locations (Biederman, 1972; Henderson & Hollingworth, 1999). Moreover in some cases individual features in the world might be more profitably linked to situations in which multiple objects occur rather than to individual objects. For example pouring, often taken as a feature of a BOTTLE,

usually occurs with a receptacle container present, and the likelihood with which pouring occurs is presumably greater when a receptacle container 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, and the binding together of language and perception in particular. Just as people are sensitive to how visual features cluster together, and how objects co-occur, so-called grounded/embodied views of cognition (Barsalou, 2008; Gallese & Lakoff, 2005; Rizzolatti & Arbib, 1998) assume tight coupling between language and non-linguistic 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/sentences involving action verbs (reading 'kick' activates motor areas involved in performing kicking actions)(Boulenger et al., 2006; Grèzes & Decety, 2001; Pulvermüller et al., 2005). However, the mechanisms underlying the grounding of language in non-linguistic systems are not well understood. One possibility is that the likelihood with which words during learning co-occur with perceptual events 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 between these possible drivers of mental animation participants completed a sentence-picture verification task in an fMRI scanner, using a 3(picture condition) x 3(language condition) design. To examine whether individual objects or knowledge of how objects interact in context drives motion processing of static images the pictures were manipulated. Consider an image of a bottle positioned higher than a glass, versus a bottle positioned lower than a glass. If learning about those objects involves knowledge that those objects interact, then simulating the path of liquid falling from the bottle should only occur when a glass is presented below it; so MT+ activation should be greater when the bottle is in the functionally congruent position in which to interact with the glass compared with when the same objects are present, but in a position where pouring is not relevant for that interaction. In contrast, if knowledge of objects drives motion processing, then there should be no difference in MT+ activation when a bottle is positioned higher or lower than a 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 where those objects are positioned relatively (Coventry & Garrod, 2004). *The bottle is over the glass* is judged as being more appropriate to describe 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 missing the glass (Coventry & Garrod, 2004; Coventry et al., 2010). Hence for static images involving containers

beginning to pour liquids/objects towards other containers we predicted that spatial language judgments (e.g. *Is the bottle over the glass?)* would require motion processing of those static images in order to establish whether the bottle is indeed over the glass. In contrast, 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.

Additionally we were also interested to establish whether language mediates premotor/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 glass, but simulating the action of pouring. It has been shown that viewing objects automatically potentiates actions towards those objects (see for example Grèzes et al., 2003; Tucker & Ellis, 1998). We wondered whether such motor activations may be dependent on functional relations between objects (consistent with data from patients with visual extinction; Humphreys & Riddoch, 2007), and further, if the relevance of the action cued by language might affect the extent of such activation patterns.

Method

Participants

Twelve healthy, right handed, native English speakers (9 male, 3 female; 21-45 yrs.; mean=30.25 years) recruited from the city of Bremen participated in the experiment. All participants grew up in an English-speaking country with English as their first language, and had only recently moved to Germany.

Prior to recruiting participants, the study was given full ethical scrutiny and subsequent approval in accordance with international standards.

Experimental procedure and materials

Participants took part in a sentence-picture verification task in an fMRI scanner. The task was to judge whether sentences of the form <The NOUN1 is TERM the NOUN2> were a true or false description of a picture that followed (see Figure 1a). Sentences contained two different sets of closed class terms across three conditions: two spatial preposition conditions (Vertical Prepositions (VP)'over/under/above/below' and Proximity Prepositions (PP)'near/far') and a comparative adjective (CA) ('bigger/smaller') condition. Pictures were of three types (Figure 1b; see also Figure S1). In two associated object conditions, two objects (NOUN1, NOUN2) that frequently co-occur together were shown either placed in their functionally congruent positions (e.g., a packet higher than a pan, hereafter FC; Figure 1b, left panel) or in a functionally incongruent position (e.g., a pan higher than a packet, hereafter FI; Figure 1b, middle panel). In a third condition objects were presented that did not usually co-occur and did not interact functionally (e.g., a pepper higher than a telephone, hereafter Non-functional, NF; Figure 1b, right panel). Both the FC and FI pictures showed a container with falling objects protruding from its spout/edge at various angles (Figure 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 9 (3 picture x 3 language) conditions. These trials were matched for object position (Figure S1 D.), mirroring (Figure S1 E.), and object flow position (for

the FC and FI conditions). The sentences were manipulated in noun order and preposition direction (The bottle is over the glass; The glass is under the bottle; The glass is over the bottle, etc.) ensuring that 8 out of every 24 trials were false descriptions (so that participants would attend throughout the task).

After giving written consent, participants were instructed that the task was to indicate by button press as quickly as possible whether a picture was correctly described by the preceding sentence. The inter trial interval was jittered with a range of 0.65-1.36 seconds. A practice run comprised 8 (random) trials from the main experiment. Data for baseline and sentence-picture verification trials were acquired in a single continuous scanning run (Figure S2).

The order of the conditions in the sentence-picture verification task was determined by a pseudo-randomized non-stationary probabilistic design (Friston, 2000; see Figure S2). A simple attentional task comprising geometric objects matched in size and color was recorded as a baseline – participants were instructed to press any button when an object appeared on the screen (see Figure S1 C). In addition we used a localizer scan to individually define functional regions of interest for MT+. The localizer paradigm consisted of squares floating into the screen (radial flow field with size change) or remaining static (static field) consistent with localisers used previously (Tootell et al., 1995).

FMRI recording and analysis

Functional BOLD data were acquired with a 3T SIEMENS Magnetom Allegra™ head scanner using a whole head, local gradient coil (SIEMENS, Erlangen, Germany). Functional images were collected with an Echo Planar Imaging (EPI) sequence

during the sentence-picture-verification task, baseline task, and the localizer task developed to define functional ROIs for MT+ (TR 2.06 sec, TE 30 msec.). 38 slices of 3 mm^3 voxels were acquired over the whole cerebral cortex. High resolution T1 weighted structural scans were acquired as a set of 160 contiguous saggital slices (1 x1x1 mm voxels, MPRAGE TR=2.3 secs, TE=4.38 msec., 256x256, FA 8^0).

SPM5, MARSBAR and STATISTICA were used for analysis. After correcting for slice acquisition order and head motion, the functional data was co-registered with the structural scan. Normalizing parameters were obtained using the unified segmentation approach and normalized the functional data onto the MNI space. Finally the normalized data were smoothed with a Gaussian kernel of 9 mm (FWHM). 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, Dukelow, & Verstaten, 2001). We calculated percent-signal-change for all experimental conditions as computed by MARSBAR for MT+ in both hemispheres. The percent-signal-change data was incorporated into a repeated measures sentence type x picture type ANOVA.

We also investigated whether areas other than MT+ show consistent changes in activation in our 3x3(sentence type x picture type) design using a mass-univariate approach. To do this, we created contrast images for the nine (one per cell of the design) conditions against baseline and incorporated these images into a second level between subjects 3x3 ANOVA (N=12). We then tested for main effects of picture and language and for an picture x language interaction (all $p_{(FDR)}$<.05, k=20) and compared single conditions for the cluster peaks. Successively, we created an

overlay with the activation maps for the main effects of language and picture (red in Figure 3) areas at the temporo-parieto-occipital junction that showed significant activations in the MT+ localizer ($p_{(uncorr)}$<.05, shown in green in Figure 3).

Results

FMRI analyses

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

Successfully localizing MT+ in 10 participants (see Figure 2a), we compared mean % signal change (over baseline) of MT+ activation for left and right MT+ separately using 3(sentence condition) x 3(picture condition) analyses of variance.

For left MT+ and right MT+ there were significant effects of picture condition (left hemisphere, $F(2, 18)=13.7$, $p<.0001$, partial $\eta^2=.603$; right hemisphere, $F(2, 18)=3.70$, $p<.05$, partial $\eta^2=.291$)(Figure 2b). In both hemispheres the FC picture condition was associated with significantly greater MT+ activation compared to either of the other conditions (all $p<.05$), while MT+ activation in the FI and NF conditions did not differ significantly from one another (both $p>.4$). For left MT+, there was no main effect of language condition, but there was an interaction between sentence condition and picture condition, $F(4, 36)=2.56$, $p=.05$, partial $\eta^2=.222$ (Figure 2c, left panel). In both the preposition conditions the FC pictures showed significantly greater activation than both the FI and NF pictures (all $p<.01$), with no difference between the NF and FI pictures. No differences were present

across any of the picture conditions where comparative adjective judgements were being made (all p>.05). For right MT+, this interaction was not present (F<1, Figure 2c, right panel), but there was a main effect of sentence condition, $F(2, 18) = 5.63$, p <.05, partial η^2=.385, with greater activation overall for the preposition conditions than for the comparative adjective condition. These findings were confirmed in a further MT+ hemisphere x picture x language analysis of variance, producing a reliable 3-way interaction, $F(2, 18)=3.63$, p<.05, partial η^2=.287.

We also investigated whether areas other than MT+ show consistent changes in activation in our 3x3 (sentence type x 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 $p_{(FDR)}$<.05, k=20; see Table 1). Consistent with imaging work of spatial relation processing and spatial language processing (Amorapanth et al., 2010; Damasio et al., 2001; Noordzij et al., 2008; Wallentin et al., 2005), there was increased activation in the posterior parietal cortex (R: MNI(peak): [18 -56 20]; T: 6.74; p<.001; L: MNI(peak): [-6 -58 56]; T: 5.08; p<.001) during trials with sentences containing spatial prepositions as compared to trials that included comparative adjectives. In contrast, trials with comparative adjectives induced increased activations in early and ventral visual areas (R: MNI(peak): [26 -94 6]; T: 6.51; p<.001; L: MNI(peak): [-36 -88 -10]; T: 5.93; p<.001).

There was also found a main effect of language bilaterally at the temporo-parieto-occipital junction (Figure 3), which is usually seen as the location of MT+ (Culham et al., 2001). This activation difference is attributable to increased activation during preposition conditions versus the comparative adjective condition

(R: MNI(peak): [58 -56 4]; *T*: 5.17; *p*<.001; L: MNI(peak): [-54 -52 14]; *T*: 4.54; *p*<.001). In the same comparison, activity was increased in the right supplementary motor area (MNI(peak): [14 -4 64]; *T*: 7.65; *p*<.001) and in the left premotor cortex (MNI(peak): [-48 4 50]; *T*: 2.4; *p*<.05). No differences in brain activation were present between the two spatial preposition conditions. For the main effect of picture we only found one significant cluster at the left temporo-parieto-occipital junction (MT+, see Figure 3). Activity in this area was higher in the FC condition than both the FI and NF conditions (MNI(peak): [-54 -74 0]; *T*: 4.89; *p*<.001). There was no difference between the FI and NF conditions. No voxels show a significant interaction effect after correction for multiple comparisons ($p_{(FDR)}$<.05).

As MT+ activity can be modulated by eye movements and visual attention (Dukelow et al., 2001; O'Craven et al., 1997), it is important to note that there was no main effect of picture with respect to the Frontal Eye Fields (FEF; see Table 1). Saccadic activity is generally accompanied by increased activation in FEF (Luna et al., 1998); the picture contrasts (or the language x picture x hemisphere interaction) can therefore not be explained by differences in eye movement patterns.

Behavioral data

We calculated the number of YES responses for each participant (i.e. when the sentence preceding the picture was thought to be a correct description of the picture) for the FC images and the FI images preceded by sentences with prepositions. In particular, we wanted to test whether the position of falling objects (Figure S1 E) affected responses. Previously (Coventry et al., 2010) participants drew the expected continuation of the falling objects for all four positions of objects

(see Figure S1 D.) for the two different angles of falling objects (the functional and non-functional cases, Figure S1 E.) for the FC scenes. Based on responses, we established whether participants agreed with our prior classification of scenes as functional or non-functional. For functional scenes, their drawn trajectories should have the falling objects reaching the bottom object, with trajectories missing the bottom object for non-functional scenes. High agreement was found for vertically displaced scenes (93%) but 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 as all the scenes required mental animation to establish whether the falling objects end up in the other container or not).

A 2 preposition type (vertical prepositions; *over*, *under*, *above*, *below*, proximity prepositions: *near*, *far*) x 2 picture condition (FC, FI) x 2 position of falling objects (functional, non-functional) x 2 vertical position(near, far) within participants ANOVA produced a number of main effects and interactions indicating that participants took the path of falling objects into account when making their judgements. In particular there was an interaction between the 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 were given more YES responses ($M=0.818$) than those where the falling objects were expected to miss the recipient container ($M=0.739$), $p=.04$. This was not the case for the FI scenes, where the non-functional scenes ($M=0.875$) were given more YES responses than the functional scenes ($M=0.784$), $p=.04$.

Equivalent analyses for the comparative adjectives judgements revealed no significant main effects or interactions for these terms.

We also analysed reaction times in a 3 language condition (VP, PP, CA) x 2 picture condition (FC, FI) x 2 position of falling objects (functional, non-functional) x 2 vertical position (near, far) within participants ANOVA. There was an interaction between language condition and vertical position, $F(2, 20)=6.992$, $p=.005$. Reaction times for near positions were faster than for far positions for the PP condition ($p<.001$, Bonferroni contrast), but this was not the case for either of the other two language conditions (both $p>.05$).

In summary, the behavioural data reveal that participants did consider the potential path of falling objects when considering whether pictures were described by preceding sentence with spatial prepositions in them, but not when the same pictures were preceded by sentences with comparative adjectives. Speed of response was not influenced by picture condition, or by the path 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 not only consonant 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 occurs with greater frequency 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 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 preposition conditions compared to the comparative adjective condition. In left MT+ the effect of language was not reliable, but the interaction showed an effect of picture type only when pictures were preceded by sentences containing spatial prepositions, and not when sentences contained comparative adjectives; this pattern was not found in right MT+. We take these results as strong support for the view that how 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 some recent work showing lateralized effects of language on categorical perception (Gilbert et al., 2006).

The MT+ ROI analyses results were generally supported in the whole brain analyses, with main effects of language condition and picture condition. The coarser results overall (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 et al., 2006 for discussion of limitations in anatomical averaging).

The whole brain analyses also demonstrate that language mediates motor activations when looking at pictures that follow sentences - with greater motor activations after sentences containing spatial prepositions than sentences containing comparative adjectives. In contrast we did not find any picture condition differences with respect to motor activations, suggesting that such effects may be tied to objects in the case of motor affordances rather than object interactions (i.e., 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 different terms co-occurring with the same objects is enough to switch motor simulation on and off.

Broadly, the 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 – one representing the likelihood with which perceptual objects in particular configurations and particular types of language separately co-occur with motion events (processed in the right hemisphere), and the other (left MT+) reflecting 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.

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, 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.* 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 & 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(4),* 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 Mazziotta, J. C., & Toga, A.W. (Eds.), *Brain Mapping the Disorders*. Academic Press. pp. 33–58.

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

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences U.S.A., 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(12),* 2735-2740.

Hegarty, M. (1992). Mental animation: Inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory & 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(5),* 534-547.

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

Kourti, Z. (2004). But still, it moves. *Trends in Cognitive Science, 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., Kochunov, P. V., Nickerson, D., Mikiten, S. A., & Fox, P. T. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping, 10(3),* 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(1),* 40-47.

Margolis, E., & Laurence, S. (Eds.)(1999). *Concepts.* 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(4),* 591-598.

Pullvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Science, 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.* MIT Press.

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

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience, 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., Brady, T. J., & Rosen, B. R. (1995). Visual motion after effect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature 11*, 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(3),* 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. *of Cognitive Neuroscience, 21,* 1934-1945.

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

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

Figure Legends

Figure 1. **A.** Participants completed a sentence-picture verification task in the scanner. Sentences containing prepositions (vertical prepositions: over/under/above/below; proximity prepositions: near/far) or comparative adjectives (bigger/smaller) were followed by a picture, during which participants had to respond TRUE or FALSE by depressing keys. **B.** Examples of a picture types: FC (left), FI (middle), and NF (right). The NF object pairs were objects that are not normally found together, but 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 supplementary materials).

Figure 2. **A.** MT+ localization for two representative participants (green & blue) on a rendered brain. **B.** Region of Interest (ROI) picture condition contrasts for both left and right MT+. **C.** ROI interaction between picture condition and sentence condition in left MT+ (with right MT+ shown for comparison): VP = over/under/above/below; PP = near/far; CA = bigger than/smaller than. Contrasts were computed using Bonferroni tests: NS=Non-significant ($p > .05$); *,$p < .05$, **,$p < .01$, ***,$p < .001$. Error bars show standards errors of the mean. Triangles under the graphs show conditions with % signal change values significantly greater than zero (one-sample t-tests).

Figure 3. **A.** Differences in activation for the main effect of picture type ($p_{(FDR)} < .05$, k = 20) shown in red on a rendering of the human brain. **B** & **C.** Areas that show a main effect of language ($p_{(FDR)} < .05$, k = 20) shown on two renderings. Voxels with higher activation in the spatial prepositions conditions than in the comparative adjectives conditions are shown in red, whereas the opposite contrast is shown in blue. Figures **A** & **B** include a functional between-subject ROI for MT+ based on the localizer (shown in green).

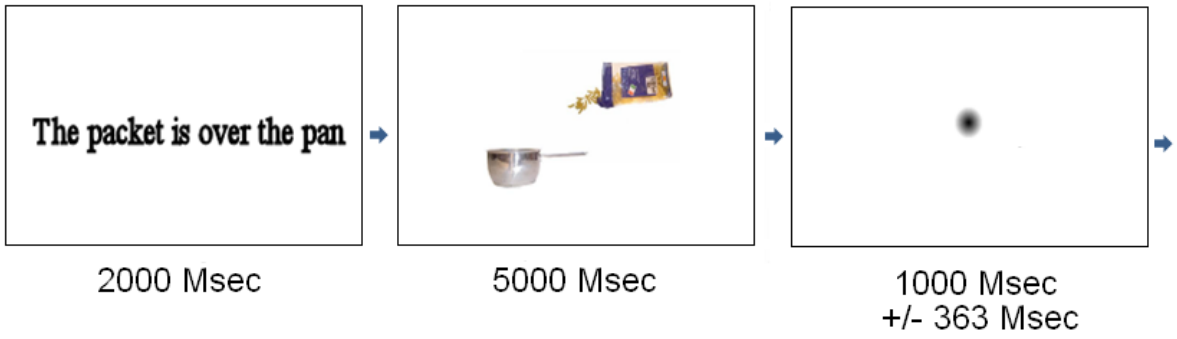
Acknowledgements

This research was supported by a Hanse-Wisschenschaftskolleg fellowship (awarded to KRC) and the Center for Advanced Imaging Bremen (BMBF01GO0506 to MH).

Author contributions: KRC conceived the study and took primary responsibility for manuscript preparation. KRC/TC designed the study, with input from TF/MH/BVC. BVC prepared the stimuli. TC ran the experiment, and took primary responsibility for neuroimaging analyses and drafting of neuroimaging results, with assistance from TF/MH/KRC. KRC analyzed the behavioral data. All authors commented on drafts of the manuscript.

Figure 1

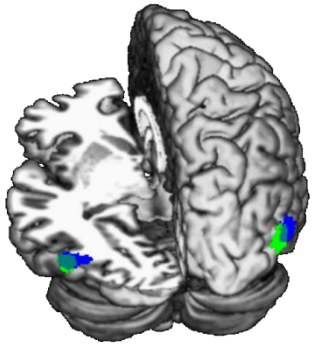
A. Trial structure



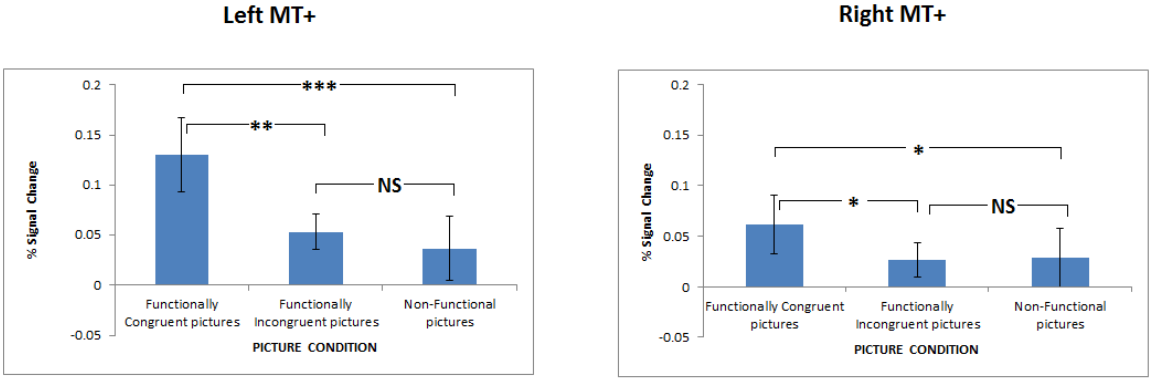
B. Picture Manipulations

Figure 2

A.



B.


Left MT+
Right MT+
% Signal Change
0.2
0.15
0.1
0.05
0
-0.05
***
**
*
NS
Functionally Congruent pictures
Functionally Incongruent pictures
Non-Functional pictures
PICTURE CONDITION

C.

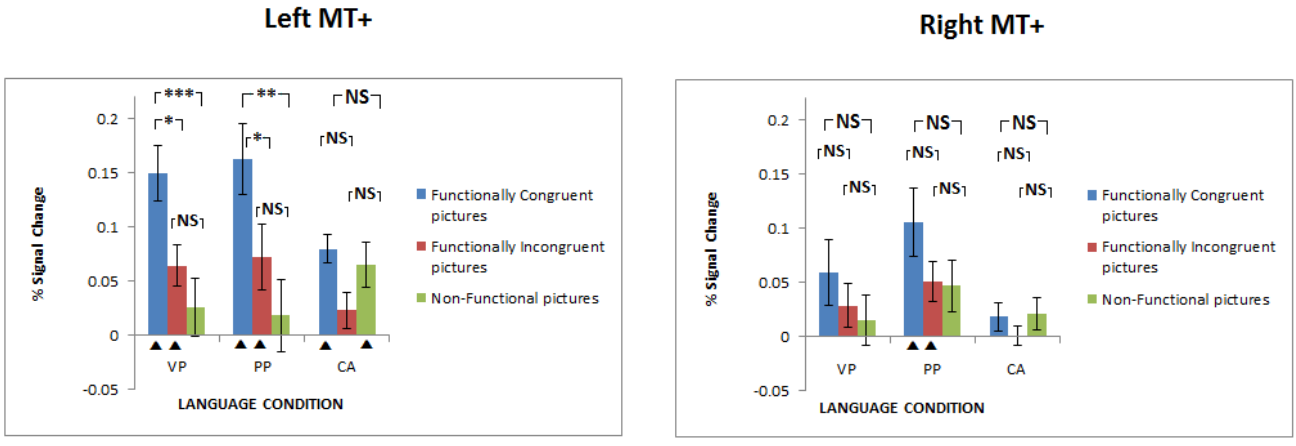

Left MT+
Right MT+
% Signal Change
0.2
0.15
0.1
0.05
0
-0.05
***
**
*
NS
VP
PP
CA
LANGUAGE CONDITION
Functionally Congruent pictures
Functionally Incongruent pictures
Non-Functional pictures

Figure 3

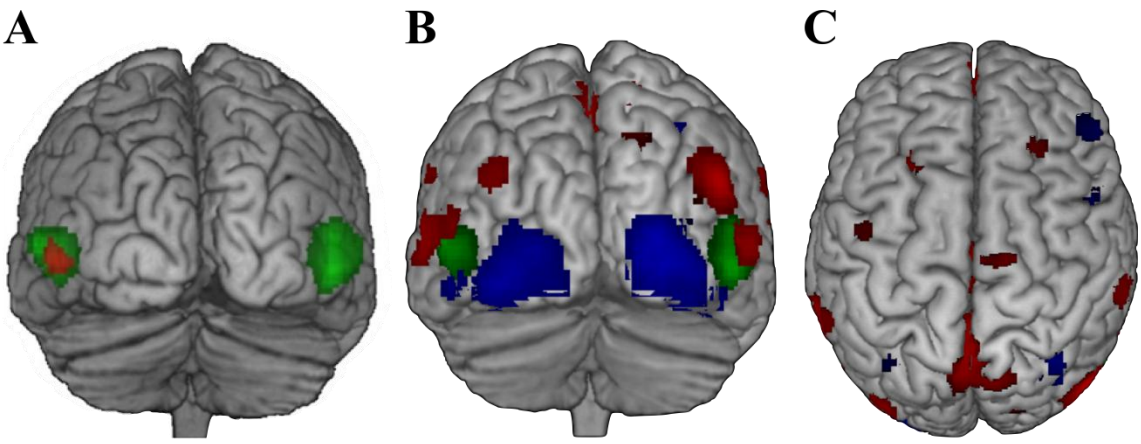


Table 1. Active clusters for the main effect of language. Anatomical labels and Brodmann Areas (BA) were assigned using anatomical labelling software (Lancaster et al., 2000) and visual inspection of the full clusters. XYZ-Coordinates of the peaks are noted in stereotaxic MNI space. F and $p_{(FDR)}$ values are based on the main effect of language. T and $p_{(uncorr)}$ are based on a contrast of the spatial preposition and comparative adjective conditions. Negative T-values mark that the given cluster responds strongest during trials that include comparative adjectives.

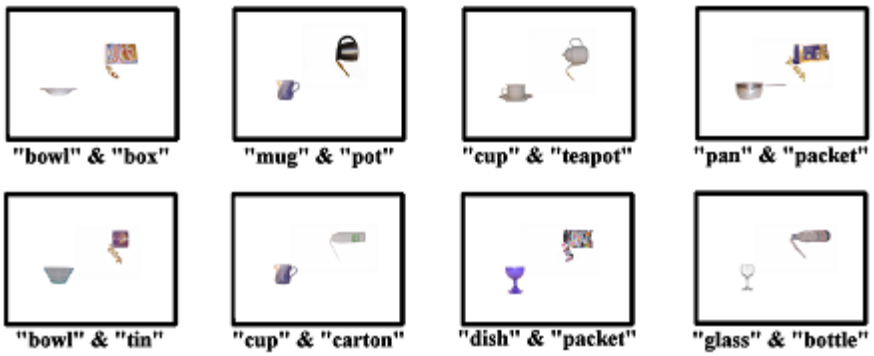
Region		**BA**	**MNI**			**Main Effect**		**SP > CO**	
			X	Y	Z	**F**	**$p_{(FDR)}$**	**T**	**$p_{(uncorr)}$**
T.-P.-O. Junction	R	22	58	-56	4	14.01	0.001	5.17	< 0.001
	L	22	-54	-52	14	12.4	0.002	4.54	< 0.001
Motor									
SMA	R	6	14	-4	64	7.65	0.021	3.5	< 0.001
Premotor	L	6	-48	4	50	7.44	0.024	2.4	0.016
Parietal									
Precuneus	R	7/31	18	-56	20	24.55	< 0.001	6.74	< 0.001
	L	7/31	-6	-58	56	17.44	< 0.001	5.08	< 0.001
Angular Gyrus	R	39	48	-74	32	22.68	< 0.001	6.49	< 0.001
Inferior Lobule	R	40	70	-32	30	11.6	0.002	3.68	< 0.001
	L	40	-64	-42	34	8.59	0.012	3.73	< 0.001
Occipital									
Middle Gyrus	R	18	26	-94	6	21.74	< 0.001	-6.51	< 0.001
	L	18	-36	-88	-10	17.81	< 0.001	-5.93	< 0.001
Superior Gyrus	L	19	-36	-84	32	10.45	0.004	4.38	< 0.001
Prefrontal									
Inferior Gyrus	R	46	52	40	14	10.27	0.005	-2.69	0.008
Medial Gyrus	R	10	6	54	-6	8.78	0.011	4.17	< 0.001
	L	10	-2	58	-4	10.2	0.005	4.35	< 0.001
Middle Gyrus	L	9	-24	26	34	8.24	0.015	4.04	< 0.001
Temporal									
Inferior Gyrus	L	21	-60	-6	-20	7.97	0.017	3.99	< 0.001
Cingulate Gyrus	L	24	-2	-16	34	12.56	0.001	4.72	< 0.001

Supplementary Online Materials (SOM-R) for

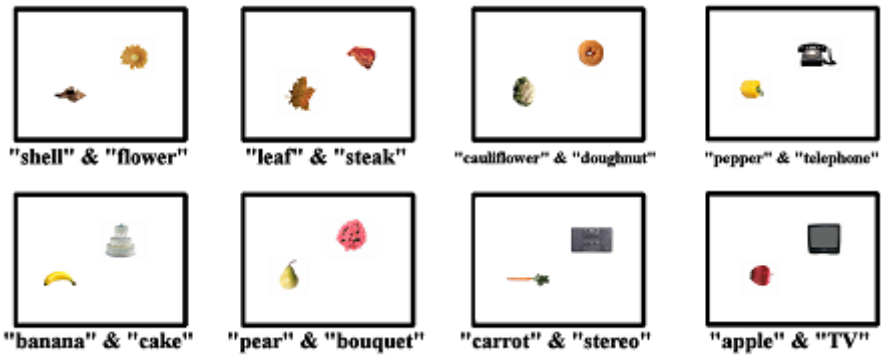
Multiple Routes to Mental Animation: Language and Functional Relations Drive Motion Processing for Static Images

Supplementary Figure S1. **A.** Shows the objects pairs used in the FC and FI conditions. **B.** Shows the objects pairs used in the NF condition. **C.** Shows the baseline objects used. **D.** Illustrates the four relative positions used (2 positions on the vertical axis: near, far; two positions on the horizontal axis: near, far). **E.** Illustrates the paths of falling objects for the FC and FI conditions. **F.** Illustrates the mirror images used: half the time one object was displaced to the right; half the time one object was displaced to the left.

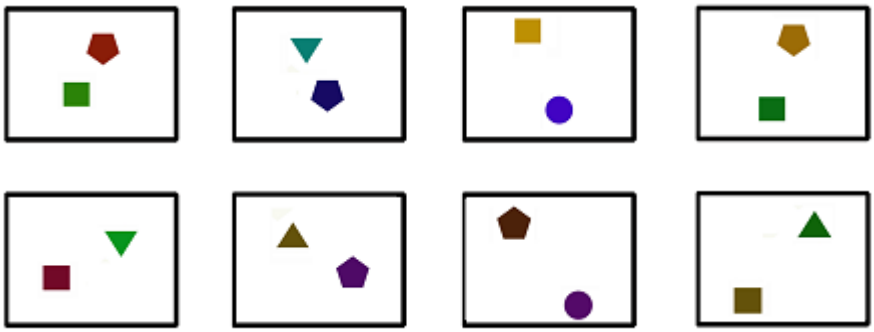
A.



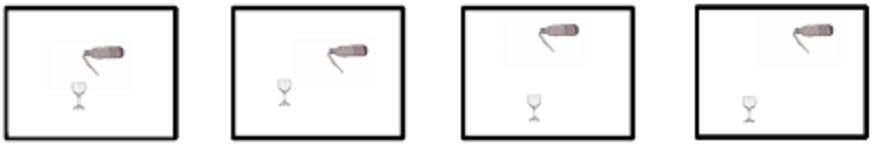
B.



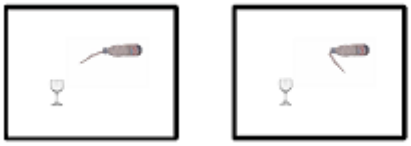
C.



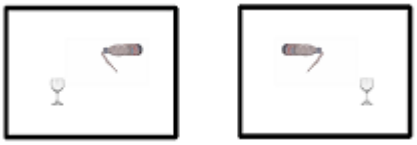
D.



E.



F.



Supplementary Figure S2. To avoid predictability and to maximize statistical power of fMRI analyses, blocks of 8 trials of a condition were intermitted by 4 trials of preceding and succeeding blocks. After an initial block of 6 baseline trials, 3 sets of 9 blocks (72 trials) were each followed by 6 trials of baseline and a break lasting 90 seconds. Block order effects were counterbalanced across participants. The top panel shows the experimental procedure for the sentence-picture verification and baseline tasks. The bottom panel shows example blocks for the pseudo-randomized non-stationary probabilistic design. 'A', 'B' and 'C' indicate single trials from different conditions.