

Handling or being the concept: An fMRI study on metonymy representations in coverbal gestures

Gina Joue^{a,b,h,*}, Linda Boven^c, Klaus Willmes^d, Vito Evola^{a,e}, Liliana R. Demenescu^b, Julius Hassemer^a, Irene Mittelberg^a, Klaus Mathiak^{b,f,g}, Frank Schneider^b, Ute Habel^b

^a Human Technology Center, RWTH Aachen University, 52056 Aachen, Germany

^b Department of Psychiatry, Psychotherapy and Psychosomatics, School of Medicine, RWTH Aachen University, 52074 Aachen, Germany

^c School of Medicine, RWTH Aachen University, 52074 Aachen, Germany

^d Section Neuropsychology, Department of Neurology, School of Medicine, RWTH Aachen University, 52074 Aachen, Germany

^e Bonn-Aachen International Center for Information Technology, 53113 Bonn, Germany

^f JARA, Translational Brain Medicine, Aachen, Germany

^g Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, 52425 Jülich, Germany

^h Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany

ARTICLE INFO

Keywords:

Multimodal communication
Coverbal gestures
Cognitive linguistics
Body-part-as-object
Transitive gestures
Pantomiming
Metonymy
Semantic attention
Abstract cognition
fMRI
TPJ

ABSTRACT

In “Two heads are better than one,” “head” stands for people and focuses the message on the intelligence of people. This is an example of figurative language through *metonymy*, where substituting a whole entity by one of its parts focuses attention on a specific aspect of the entity. Whereas metaphors, another figurative language device, are substitutions based on similarity, metonymy involves substitutions based on associations. Both are figures of speech but are also expressed in coverbal gestures during multimodal communication. The closest neuropsychological studies of metonymy in gestures have been nonlinguistic tool-use, illustrated by the classic apraxic problem of body-part-as-object (BPO, equivalent to an *internal metonymy* representation of the tool) vs. pantomimed action (*external metonymy* representation of the absent object/tool). Combining these research domains with concepts in cognitive linguistic research on gestures, we conducted an fMRI study to investigate metonymy resolution in coverbal gestures. Given the greater difficulty in developmental and apraxia studies, perhaps explained by the more complex semantic inferencing involved for external metonymy than for internal metonymy representations, we hypothesized that external metonymy resolution requires greater processing demands and that the neural resources supporting metonymy resolution would modulate regions involved in semantic processing. We found that there are indeed greater activations for external than for internal metonymy resolution in the temporoparietal junction (TPJ). This area is posterior to the lateral temporal regions recruited by metaphor processing. Effective connectivity analysis confirmed our hypothesis that metonymy resolution modulates areas implicated in semantic processing. We interpret our results in an interdisciplinary view of what metonymy in action can reveal about abstract cognition.

1. Introduction

This paper presents a study on *metonymy* in coverbal gestures, the hand movements we make as we speak. Metonymy is often defined as a “stands for” relationship: an aspect of an entity or a related entity “stands for” the entity itself. These substitutions are possible because the entities either physically coincide in space or time, are consecutive in time, or are semantically associated (known in cognitive linguistics as physical, temporal, or conceptual *contiguity*, respectively). This is in contrast to the more commonly known figure of speech, metaphors, which allow ideas to be substituted for each other based on physical or

conceptual *similarity*. There are various ways of categorizing the types of substitutions that occur in metonymy. In this paper, we focus in particular on *external metonymy* (*ext*) and *internal metonymy* (*int*) in coverbal gestures (Jakobson and Pomorska, 1983; Mittelberg and Waugh, 2014). The distinction is subtle but, as we will argue in this paper, important to understanding how we process symbolic representations; hence, we dedicate some time in explaining this distinction, which has not gained much research attention, drawing from concepts across research areas.

* Corresponding author at: Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany.
E-mail address: joue@humtec.rwth-aachen.de (G. Joue).

1.1. External metonymy (*ext*) vs. internal metonymy (*int*)

External metonymy (*ext*) refers to substitutions that are physically distinct, whereas *internal metonymy* (*int*) refers to substitutions between intrinsic parts of the same physical item. An example of *ext* in spoken or written language would be the commonly made substitution of tools or products with their producers, as in the example, “Have you read Shakespeare?” Here, “Shakespeare” refers to the works that Shakespeare has written and not to the person proper. In this example of *external metonymy*, the relationship involves two physically distinct items: (1) the person Shakespeare, and (2) the works that this person has written which are not a physically intrinsic part of himself and hence *external* to him. In contrast, a common type of *internal metonymy* is part-whole relationships such as, “Two heads are better than one,” where “heads” refers to people while focusing the message on the intelligence of people. In other words, through substitution, metonymy focuses attention on an aspect of the topic of interest.

The following elaborated example demonstrates the order of complexity, or greater semantic distance between the symbolic representation (here, words used) and the represented (the physical objects in the world that the words refer to), that can be subsumed in metonymy relationships. When the waiter at a restaurant asks a customer, “Would you like another cup?” there could be multiple intended meanings given the context and background knowledge. The different interpretations that can be drawn can be graded on degrees of inference complexity. If the guest had a pot of coffee on his table and just accidentally broke his cup, the waiter is literally asking the guest whether he would like him to bring another empty cup (“cup” literally refers to the object itself). Successful interpretation of the utterance requires a representation of an absent cup and is what we call first-order inference. If the guest did not have this coffee pot and broke his cup, then the waiter is asking whether he should bring a new cup that is also filled with coffee. If the guest had drunk all his coffee and thus emptied his coffee cup, the “cup” also does not refer to another cup but to “more coffee” and is an example of the commonly used *ext* relationship *CONTAINER FOR CONTENT*, where the content or target referent is *external* to, or not an intrinsic part of, the symbol/word used. We can consider the last two interpretations as examples of second-order inference as they not only require a representation of the absent functional cup but also of the content that is usually in such a cup (object + association). Thus, metonymy, like metaphor, is a mechanism for creating representations and, inversely, making inferences (Barcelona, 2009; Panther and Thornburg, 2003). It is a powerful mechanism that we commonly use to speak or write in shorthand.

1.2. Metonymy is about understanding associations, regardless of modality

Metonymy is not just a mechanism of spoken or written language but also occurs in other visual or auditory media. A scene of a messy array of footprints can indicate a previous struggle. The sound of silence following sounds of a struggle, ended by the sound of a gun being fired, can imply the end of a struggle. Metonymy also allows us to only have to partially represent what we are conveying through the hand gestures we make as we speak. Because gestures will physically ever be able to only partially represent what is intended and can only offer components of a message that the interpreter must synthesize (McNeill, 1992), all communicative gestures reflect metonymy (for detailed discussions, see Mittelberg and Waugh, 2009; Littlemore, 2015). However, in order to create such a synthesis, the interpreter must infer what aspect of the hand movements is relevant, (e.g. Which part of the hand? The shape or the movement? Or the trace left by the hand movement?). That is, understanding coverbal gestures requires attention to be focused on intended aspects, which in turn must be inferred.

Returning to our restaurant example, the waiter might have made an *iconic* gesture, that is, the hand shaped to physically resemble a cup (e.g. the palm up and fingers upwards slightly bent). When referring to

the cup, this gesture reflects *int*, as it is iconically representing the perceived essential form features of the body of the cup itself (and hence the gesture represents a part *internal* to the object). A first-order inference must be made to interpret the gesture as taking on a form that physically resembles the signified, or represented object. When referring to coffee, however, this gesture reflects *ext*, as it is representing the cup, that is, a mass external to the structure of the coffee. It not only draws on the first-order inference that the hand shape represents a coffee cup, but also on the associative relations between cups that are used to hold coffee and the context.

The waiter might also have made a hand gesture as if holding the handle on a mug, with index finger slightly curled, thumb straight and other fingers curled more tightly. Whether the gesture is intended to refer to a cup or the coffee, it is an example of *ext* as the hand holding the cup is not an intrinsic property of the cup nor the coffee. It requires second-order inference: it requires inferencing that the gesture represents an action based on the affordances of a cup, an object which also needs to be imagined (Table 1).

1.3. Metonymy as representation in gesture: grasping or being

The importance of the *int* and *ext* distinction becomes more apparent when we turn to aphasia and developmental research. Although the concept of metonymy in the theoretical linguistic analysis of coverbal gestures has only been developed relatively recently (Mittelberg and Waugh, 2009), it is analogous to the classic problem, in ideomotor apraxia and in developmental psychology, of *body-part-as-object* (*BPO*; *int*), where a body part is used to represent an object, such as a tool, versus *pantomiming* with an imagined tool or *object substitution* (*ext*). Ideomotor apraxia is a neurological condition resulting from brain damage, characterized by difficulties with the ideation (determining what action/interaction is appropriate), action planning, and execution of conceptually learned actions (e.g. waving goodbye or combing one's hair). Although this paper reports on healthy adult individuals, these areas of research provide useful insights in the little-researched area of metonymy. Effectively, research in ideomotor apraxia and in developmental psychology suggest that with healthy development and the healthy mature brain, pantomiming (*ext*) occurs more commonly than *BPO* actions (*int*). The younger, less developed brain and the compromised brain appear to tend toward more *BPO* actions. If we can consider actions from younger minds and apraxia as possibly favoring simpler actions, we could infer that *BPO* (*int*) actions are cognitively less demanding than pantomiming actions (*ext*).

1.3.1. Pantomiming reflects a better symbolic grasp: apraxia

When asked to demonstrate daily tool-oriented actions in the physical absence of the tool (e.g. brushing teeth), healthy adults will tend to pantomime the action with the imagined tool (e.g. toothbrush, *ext*). In contrast, adults with ideomotor apraxia will tend to perform *BPO* (*int*) over pantomiming (*ext*) (Goodglass and Kaplan, 1963). More crucially, patients have difficulties performing pantomimes even when insistently re-instructed (Raymer et al., 1997), although this seems to be in patients with left-lateralized lesions (see Binkofski and Buxbaum, 2013). However, they have no problems performing these familiar actions in real contexts (De Renzi et al., 1980), suggesting that the problem might be less of motor execution than motor association or planning. *BPO* (*int*) arguably provides a more concrete iconic representation of the absent tool, whereas pantomiming with an imagined object (*ext*) requires both representation of an absent object and of the associated action (Frey, 2008). Some support for this representational view comes from Clark et al. (1994), who found that the kinetics of apraxic movements improves with increased context provided by the presence of an object and/or tool. Although most of these studies have focused on manual gesture production as a diagnostic tool for ideomotor apraxia, correct recognition of pantomimed movements is also worsened in patients with ideomotor apraxia with more severe symptoms,

suggesting that their difficulties in pantomiming are seated in problems in symbolic representation (McDonald et al., 1994).

Neuroimaging studies on praxis representations in adults with no neurologically related motor problems suggest that metonymy resolution in gestures might involve areas in the inferior frontal gyrus (IFG), lateral temporal cortices, and inferior parietal areas – all areas that, as we will see in the next section, also happen to be reported in semantic processing. For example, the IFG and lateral temporal cortices, along with the inferior parietal lobe (IPL), have been consistently reported in the context of representations required for pantomiming tool-use actions (*ext*) (Frey, 2008). The left posterior superior and middle temporal gyri (pSTG and pMTG, respectively) have been reported in naming the actions associated with a tool rather than just naming the tool itself (Martin et al., 1995). Imagining the pantomiming of tools (*ext*) compared to BPO (*int*) gestures also elicits greater activity in the right pars opercularis of the IFG (IFGop) (Higuchi et al., 2007) (although note that executing BPO actions with the right hand elicits greater activity in the right SMG compared to right-handed pantomiming: Ohgami et al., 2004). In viewing these results with those from developmental studies of actions in children, we see an even stronger link between action and symbolic representation.

1.3.2. From being to grasping: developmental research

True symbolic representation in child development has been argued to parallel object substitution during pretend play (most influentially, e.g. see Piaget, 1962; Vygotsky, 1976). Like patients with apraxia, young children have difficulty with pantomiming (*ext*). Children under 3–4 years old will predominantly perform BPO (*int*) actions and have difficulties pretending with an imagined object (*ext*) (Mitchell and Clark, see also (2015)). However, as they develop normally, their actions will more often involve imagined or substituted objects (Elder and Pederson, 1978; Boyatzis and Watson, 1993; O'Reilly, 1995), which can become progressively less similar iconically (Elder and Pederson, 1978) and functionally (Pederson et al., 1981) to the represented object. The developmental transition towards pantomiming actions (*ext*) seems to correlate with Theory of Mind and other symbolic abilities, which might correspond with age (Dick et al., 2005; Suddendorf et al., 1999).

Other studies have also found correlations between object substitution and language. This is an important association if we consider language as a complex system of symbolic representations and the development of linguistic cognition as an indication of the ability to handle symbolic representations in general. For example, 13-month-old infants who substituted with objects that were incongruent to the pretend action (e.g. drinking from a toy car) had larger expressive vocabularies. More frequent substitutions with featureless, hence more abstract, objects (e.g. a wooden block) also correlated with larger receptive vocabulary (Bates et al., 1980). Similarly, pantomime (*ext*) comprehension at 24 months appears to be correlated with receptive language abilities, such as word meaning and semantic relationships, at 36 months (O'Reilly et al., 1997). Piaget (1962) referred to this development as an internalization of representations and the ability to differentiate the symbol (or *signifier* in the semiotic system) and the represented (*signified*). This development permits greater flexibility and transformations between more distant spatiotemporal relationships (Piaget and Inhelder, 1971; Werner and Kaplan, 1963).

Given the apraxia and developmental behavioral research indicating greater difficulty in pantomiming gestures (*ext*), as reviewed above, we expected that coverbal gestures reflecting *ext* representations would require greater processing demands than gestures with *int* representations. Increased cognitive demands should correspond to greater activity in semantic attention networks responsible for reorienting attention to relevant aspects of a stimulus. That is, the brain regions involved in metonymy resolution should modulate those required for retrieving semantic associations.

1.4. Possible neural networks for metonymy resolution in communication

To our knowledge, only one neuroimaging study has investigated linguistic metonymy. This study found that metonymy resolution in written language, compared to literal counterparts, elicits a predominantly left-lateralized network including the IFG and MTG (Rapp et al., 2011). Unsurprisingly, the IFG, STG and MTG have been consistently reported in semantic processing, whether nonlinguistic or linguistic (for meta-analyses, see Cabeza and Nyberg (2000); Vigneau et al. (2006) as well as for abstract concepts (meta analysis: Wang et al., 2010), such as metaphors (for a meta-analysis, see Rapp et al., 2012). The MTG and STG have been suggested to play a role in semantic integration in multimodal communication (Straube et al., 2008, 2011; Kircher et al., 2009; Willems et al., 2009; Holle et al., 2010; Dick et al., 2014). Studies have also found the IFG to be involved in multimodal semantic processing (Chee et al., 1999; Buckner et al., 2000), including coverbal gestures (Dick et al., 2009), inferential predictions (Virtue et al., 2006; Chow et al., 2008), and general selection of semantically-relevant information (Wagner et al., 2001; Badre et al., 2005), but only when prior semantic communicative association exists (Capek et al., 2010; Holle et al., 2008). Both left IFG and left MTG have also been found to be involved in processing abstract semantic information whether in speech or in gestures (Straube et al., 2013). Unsurprisingly, IFG is intrinsically connected to the STG/STS (Turken and Dronkers, 2011) via the temporal-parietal junction (TPJ) (Hickok and Poeppel, 2007; Scott and Johnsrude, 2003) and has been reported to modulate the left pSTS in speech–gesture congruence assessment (Willems et al., 2009) and to be a part of a network responsible for speech comprehension (Hickok and Poeppel, 2007) and semantic predictability (Obleser et al., 2007).

The TPJ, in turn, has been implicated in a multitude of roles including inferences in social cognition (Van Overwalle and Mariën, 2016), successful speech comprehension under difficult acoustic conditions regardless of semantic predictability (Obleser et al., 2007), and integrating semantic information across modalities especially when semantic judgments are difficult (for a meta-analysis, see Binder et al., 2009). The TPJ has been proposed to play an attentional role that helps filter perceptual input (Serences et al., 2005) that is relevant (Corbetta et al., 2000; Downar et al., 2000; Indovina and Macaluso, 2007). This is consistent with patient studies that suggest that lesions in the right TPJ are related to neglect across sensory modalities (Pavani et al., 2003).

In summary, many studies have implicated the IFG, STG/MTG, parietal areas, and the TPJ, in understanding linguistic semantics, coverbal gesture perception, and in praxic representations. Coverbal gestures tie all these three research domains. Studies on apraxia, pretend play, and linguistic metonymy point to how metonymy resolution in hand actions can reflect symbolic cognition, and that pantomimed/*ext* actions might require greater resource demands, in more semantic processing or increased semantic attention, than BPO/*int* actions. This resolution possibly involves the IFG, lateral temporal cortices, or inferior parietal areas such as the TPJ. In order to test these hypotheses, we conducted an event-related fMRI study of multimodal communication understanding involving metaphors in speech and gestures. We hypothesized that the neural networks involved in metonymy resolution would modulate networks for semantic understanding. We also hypothesized that metonymy is an inference device distinct from metaphor processing. In multimodal communication, semantics is conveyed verbally and/or gesturally. We therefore distinguished between utterances that involved literal sentences and iconic gestures as well as metaphors expressed in speech only (*monomodal metaphoric gesture*) or in both speech and gesture (*multimodal metaphor*, Müller and Cienki, 2009).

2. Materials and methods

2.1. Subjects

The data reported here were from 32 healthy volunteers aged 18–31 years old (11 female, mean age 24.5 years, SD 3.5 years). All participants were native or near-native German speakers (acquired German before 5 years old), right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971), and had no history of neurologic problems as reported by the participants or psychiatric issues according to SCID screening (Wittchen et al., 1997) using the SKIDPIT Light (Demal, 1999). The study protocol was approved by the local ethics committee of the Medical Faculty at the RWTH Aachen University, and all volunteers gave written informed consent. They were paid €10 for their participation in this fMRI study.

2.2. Stimuli

Stimuli for the study were 140 3–4-sec audiovisual videos (Fig. 1). Each video showed the frontal upper bust view of a male actor uttering one German sentence and producing a corresponding coverbal gesture or a grooming action. The videos were framed to maximize the focus on the gesture stroke, the meaningful content phase of the gesture (Cornejo et al., 2009; McNeill, 2005). As our aim was not to focus on audiovisual/speech-gesture integration but on the mechanisms of metonymy and the difference between metonymy from other figurative language mechanisms like metaphor in the context of multimodal communication, we did not have a speech-only condition and do not test simple effects (i.e. experimental conditions against an implicit baseline). Speech-only stimuli would not take into account audiovisual integration and nonlinear additive effects of constituent sensory modalities. We only compare conditions that have both audio and visual modalities, namely, speech with a manual gesture. All communicative manual gestures here inherently reflect either internal metonymy or external metonymy. Thus, to isolate metonymy without metaphoricity, we also had non-metaphoric and literal speech-gesture combinations, labeled as “iconic” (ico). “Iconic” gestures bear resemblance to what is represented (e.g. flat palm-up hand to represent a flat piece of paper; McNeill, 1992).

Videos could be categorized according to speech metaphoricity (literal or metaphoric) and communicativity of gestures (communicative or non-communicative) (Figs. 1 and 2). Videos with communicative gestures were further categorized according to metaphoricity (iconic or non-metaphoric vs. metaphoric, described shortly) and metonymy type (ext or int; see Section 1.2 for explanation). Whether a

gesture is interpreted literally or metaphorically depends on the affiliate speech content (e.g. Mittelberg and Waugh, 2009, 2014). We differentiated metaphors that were expressed monomodally in the gesture accompanied by a literal spoken utterance (mono), from metaphors expressed verbally while the gesture provided redundant or iconic information (multi). We included non-communicative, or grooming actions (gro), such as, fixing hair, scratching, rubbing nose, and fixing shirt. As grooming actions do not refer to or represent content, they cannot reflect int or ext. Therefore, our factors of video/gesture category (ico, multi, mono, gro) and metonymy type (int, ext) do not fully cross. There were seven conditions in total (Fig. 2). Metaphors will only be compared with metonymy as different types of figurative language. In-depth discussion of metaphors is out of the scope of the current paper.

Cleanly balancing sentence and hand action tokens across all video categories was not possible without introducing sentence-gesture combinations that were difficult to categorize or unnatural. We followed a rigorous procedure to minimize differences in characteristics of sentences and gestures among video categories, i.e. to balance gesture and sentence complexity, across all video types and to validate videos. Stimuli were constructed so that, when possible, the same gesture was kept for as many utterances and gesture types as possible: 90% of the same hand shapes, or gestures, were distributed evenly ($\pm 5\%$) across video categories. Because studies have found that meaningful gestures tend to be produced by the dominant hand (Kimura, 1973; Lausberg et al., 2007), all video categories had a balanced number of right-handed and bimanual gestures.

In a pilot study, 18 naive native German speakers judged the naturalness and emotional neutrality of 246 of the recorded videos on a 7-point scale. We expected that gestures would bias the imageability of the abstractness of the video content (Wiemer-Hastings and Xu, 2005). Because the metaphoric mapping for monomodal metaphoric videos is instantiated in the gestures and because gestures are less well-formed than textual media, we did not ask participants to judge the familiarity, conventionality or imageability of the videos.

Three expert annotators also independently coded the videos according to video type (gesture type and metaphoricity of the speech), to corroborate each assigned video category. Videos which were rated as non-neutral or had the highest number of disagreements among annotators were excluded. More specifically, the best ranked videos were selected according to a summary score based on the following criteria (the most heavily weighted listed first): lowest total number of disagreements among all experts across all categories, lowest maximum number of disagreements among experts for any given coding category, emotional neutrality, and the greatest number of videos across different

Table 1
Inference complexity across metonymy types in verbal and gesture modalities, by example.

modality		expression		signified	
				cup	coffee
verbal		“cup”		int	ext
gesture		cupped hand		int	ext
gesture		pantomiming holding handle of mug		int	ext

The degree of inference complexity depends on what is expressed (expression) in what modality to refer to what (signified). It increases from internal (int) to external (ext) metonymy. For example, a waiter asking, “Would you like another **cup**?” when referring to **coffee**, which is external (ext) to the “cup” expressed requires the interpreter to additionally infer that the waiter is not literally referring to a **cup** but to **coffee**, which is not expressed at all, compared to if the waiter said “cup” to literally mean **cup**.



Fig. 1. Example gesture-utterance combinations for the gestures depicted in the subfigures. Utterance used for ico.int with gesture (a) above: “Die Korken gingen auf dem Wasser auf und ab.” (“The corks bobbed in the water.”), i.e. fists representing the corks themselves. Utterance used for ico.ext with gesture (b): “Die Bälle des Jongleurs waren schwierig zu sehen.” (“The juggler’s balls were hard to see.”), palm-up open-hand representing the hands on which the juggler’s balls rest, hence external to the signified. Utterance used for multi.int with gesture (b), multi.int with gesture (a), and gro with gesture (c): “Ich muss die Vor- und Nachteile abwägen.” (“I need to weigh the pros and cons.”). Utterance used for mono.ext with gesture (b) and mono.int with gesture (a): “Ich kann mich nicht entscheiden.” (“I can’t decide what to do.”). Note that it was not possible to use the same utterance across all conditions without violating naturalness of gesture-utterance combinations and increasing ambiguity of internal/external metonymy categorization. Naturalness of gesture-utterance combinations was validated by a pilot study of naive participants and 3 expert annotators. Video categories were validated by 3 expert annotators. Gesture-utterance combinations were chosen foremost based on greatest inter-rater reliability of categories, naturalness, and the criterion that the same gesture be used across as many conditions as possible (see main text for details). The examples presented here represent the case when a single gesture spanned a maximum number of conditions: the same gesture usually covered only ext conditions or only int conditions. Utterances for gro condition were balanced across utterances used for ico, multi and mono.

categories using the same gestures. Each video was also qualitatively reviewed and a handful of videos were re-recorded or newly recorded. New videos were then reviewed and recoded by three annotators. For the final set of 140 videos, with 20 videos for each of the seven conditions, the assigned video categories had an overall $\kappa = 0.875$, indicating an excellent level of agreement among annotators. The average rating for the communicative gesture videos in this set was 3.6 ($SD = 1.5$). Naturalness and emotional ratings did not differ between metonymy types (p ’s > 0.4).

Each video was trimmed to 400 ms (10 frames) before mouth opening (manual movements always occurred later) and 400 ms after the last hand or mouth movement. Audio of the videos was equalized between left and right channels. Videos varied in duration. The average duration of the videos was 3.60 s ($SD = 0.47$ s), amounting to an average of 5.19 words ($SD = 1.69$). There was no difference between metonymy types in video lengths, average word count per utterance or per T-unit (96.7% of the utterances had only 1 T-unit), phonological complexity, or phonetic complexity (p ’s > 0.7), nor in cumulative lexical frequencies or syntactic complexity (χ^2 tests, p ’s > 0.2). Phonological complexity was assessed by the number of word-initial late-emerging consonants per utterance (Shriberg, 1993). Phonetic complexity was assessed by calculating the Index of Phonetic Complexity (IPC; Jakielski, 1998) for each utterance. The IPC is a composite score based on 8 different production features: consonants by place (penalty for dorsals), consonants by manner (penalty for fricatives, affricates, liquids), vowels by class (penalty for rhotics), words by shape (penalty for consonant-final words), number of syllables (penalty for trisyllabic words and greater number of syllables), consonant series (penalty for consonant clusters), and cluster types (penalty for consonant clusters

with different places of articulation). Cumulative lexical frequencies for utterances were estimated with the SUBTLEX-DE lexical frequency for German, based on a large corpus (over 25 million words) of subtitles from films and television series, and shown to be a better indicator of lexical frequency than Celex, dlexDB, or Google Ngram, as well as, a better predictor of lexical decision RTs (Brysbaert et al., 2011). Syntactic complexity was scored according to the modified D-level scale (Covington et al., 2006), based on developmental research in language acquisition (Rosenberg and Abbeduto, 1987).

2.3. Experimental design and procedure

Participants were informed that the study was about how people communicate (Fig. 3). They were verbally instructed to simply watch the short videos and answer a yes/no question (“Was the topic of the video concrete?”) after each video. Although, arguably, the question could be answered by speech alone, we monitored that they were watching the videos through an eye-tracking camera. We intentionally did not mention gestures to participants as we were interested in how coverbal gestures influenced responses without focusing attention on them, as in natural communication settings.

Responses were recorded as button presses: right index finger for “yes” and right middle finger for “no.” Participants were told that there was no correct answer to this question but to use the guideline that a physical object or something “touchable” was “concrete”, such as a table, and what was not physically touchable was “abstract”, such as love. Further elaborations were not made as we were interested in whether higher orders of inference, elicited by the different metonymy types and their interaction with metaphors, would affect people’s

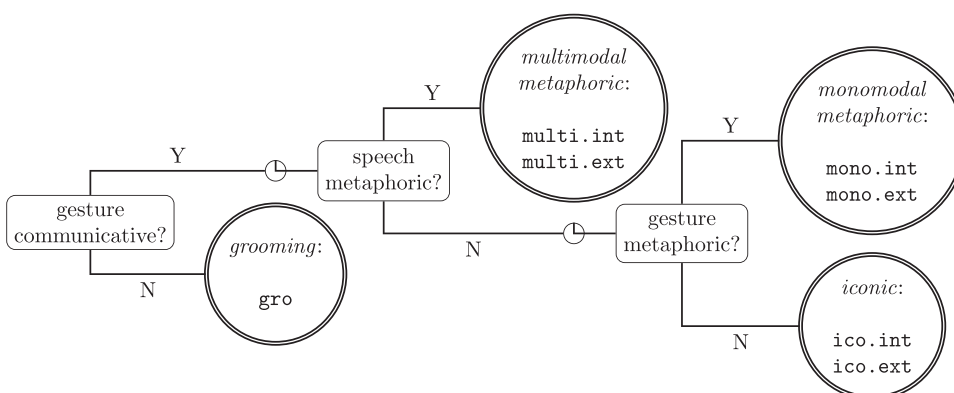


Fig. 2. The videos we used in this study included two types of communicative coverbal gestures, iconic (ico) and metaphoric gestures (multi and mono), and non-communicative hand actions (gro for “grooming”). Communicative coverbal gestures could be further divided into representations reflecting internal metonymy (int) or external metonymy (ext). Grooming gestures (gro) were included as non-communicative gestures and treated as not having a metonymy type. Videos can be characterized according to gesture communicativity (Y = yes, N = no) but also speech metaphoricity, gesture metaphoricity, and gesture metonymy type: internal metonymy (int) or external metonymy (ext). In total, we had 7 video categories: ico.int, ico.ext, multi.int, multi.ext, mono.int, mono.ext, and gro.

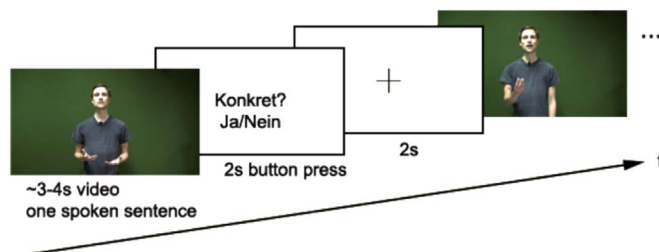


Fig. 3. Task setup. Participants viewed 3–4-s videos of an actor gesturing while saying one sentence. Following each video, they had two seconds to respond to the same yes/no question of whether the topic of the video was concrete (*Konkret? Ja/Nein*). A 2-s fixation cross then appeared before the next video played.

perception of the abstractness of a topic. Written instructions were also given. Before each video, a 2-s fixation cross was presented. After each video, which varied between 3 and 4 s, participants had 2 s to respond to the question. Halfway through the experiment, participants could take a maximum 1-min break (they could continue earlier by pressing a button with their right pinkie).

Participants practiced the task outside the scanner with four videos that were not included in the set of 140 videos used in the main study. During the main fMRI study, videos were delivered visually and aurally via Resonance Technologies' MR-compatible goggles and headphones. Audio volume was adjusted for each participant. Each of the 140 videos was presented only once in random order in an event-related design. This fMRI task lasted about 20 min.

2.4. Data acquisition

MRI scans were acquired on a 3T Siemens MAGNETOM Trio scanner with a standard 12-channel head matrix coil. Functional images of the whole brain were acquired with an EPI sequence in one session, comprising 36 axial slices in ascending order of 3.1 mm thickness with a 0.5 mm slice gap at TR = 1.92 s, TE = 30 ms, flip angle = 76°, FoV = 200 × 200 mm², and matrix size = 64 × 64. After the functional scans, a high-resolution T1-weighted MPRAGE structural scan (resolution isotropic 1 mm) of each participant was acquired as well (TR = 1.9 s, TE = 2.52 ms, flip angle = 9°, FoV = 250 × 250 mm², matrix size = 256 × 256).

2.5. Behavioral data

Judgment of abstractness/concreteness of concepts is known to be influenced by a complex interaction of numerous factors (Wiemer-Hastings and Xu, 2005; Ghio et al., 2013). We tested whether metonymy type influenced participants' judgments of concreteness in a one-factorial repeated measures ANOVA with Type III sums of squares on the arcsin transformation of the binary abstract/concrete responses (to normalize the distribution of binomial random variables). ANOVAs were done with the *aov_car* wrapper function of the CRAN R library *afex*.

In case differences in metonymy type resolution differed in processing time, we analyzed response times (RTs). We also fitted ex-Gaussian distributions to the RTs to address the right-skewedness of RT distributions. This is reported in the Supplementary Information. Briefly, the ex-Gaussian distribution models RTs as a convolution of a Gaussian and an exponential function, where the exponential function represents the right tail of the distribution (Hohle, 1965; Ratcliff and Murdock, 1976). Ex-Gaussian distributions are characterized by three parameters: μ the normal mean, σ the normal standard deviation, and τ the parameter of the exponential function (see Suppl. Info for details).

2.6. fMRI analysis

2.6.1. Image preprocessing

The first three scans in each neuroimaging run were discarded to allow the signal to stabilize. Statistical parametric mapping (SPM8, Wellcome Department of Imaging Neuroscience, London, UK) was then used to correct for motion between slices, correct slice timing using the middle slice as a reference, and normalize functional and anatomical images to the Montreal Neurological Institute (MNI) standard space. Timepoints where the mean signal exceeded ± 5 standard deviations (SD) from the baseline level were automatically tagged and included later in modeling as nuisance covariates. Normalization parameters were derived from coregistering the gray matter segmented from the structural T1 image with the MNI template and the mean EPI functional image. The normalization parameters from the mean EPI image were then applied to all the EPI images. Normalized EPI images were smoothed by an 8-mm FWHM Gaussian kernel. The time series in each voxel were high-pass filtered at 128 s to remove low-frequency noise and slow drifts.

2.6.2. Whole brain analysis

The onsets of the video events were convolved with a canonical haemodynamic response function (HRF) and included in a general linear model for each participant in a mini-epoch analysis, accounting for the varying lengths of the videos and also acting as the jitter to reduce correlation between regressors. Grooming gestures were only included in the model to account for video events with grooming gestures, rather than including them with the baseline, but will not be further discussed in this paper as they are not relevant to metonymy representation. Because images were acquired during presentation of the instruction and during the midway break, which both varied in length, these were included in the model to regress out non-task related activity. This resulted in the following nine regressors in the general linear model used in our whole brain analysis (see also Fig. 2): (1–2) *ico.int* / *ico.ext*, (3–4) *multi.int* / *multi.ext*, (5–6) *mono.int* / *mono.ext*, (7) *gro*, (8) instructions, and (9) rest. For each participant, outlier spikes in the signal as identified during preprocessing, were also included in the model as nuisance covariates. Six nuisance covariates for residual movement (three translational and three rotational), determined during spatial realignment, were also included as nuisance regressors to account for linear residual trends. Temporal autocorrelations were corrected using AR(1), an autocorrelation correction model where serial correlations are modeled by regressing out the variance explained by the immediately prior scan. No global normalization was done.

Due to the inseparability of metonymy type and metaphoricity in multimodal communication, a direct comparison of metonymy with metaphoric processing was not possible. We tested whether *ext* resolution requires greater processing demands than *int* representations (*ext* > *int*), by averaging across metaphoricity (i.e. *ico.ext* + *multi.ext* + *mono.ext* > *ico.int* + *multi.int* + *mono.int*). Similarly, to nonetheless also address the likelihood that metonymy and metaphor recruit different neural resources despite this intrinsic limitation, we compared the brain regions recruited by *ext* > *int* with those involved in supramodal metaphoric processing, based on results from a separate analysis reported elsewhere (Joue et al., in press, which reports on metaphors in much more depth that is not relevant to the present discussion). This supramodal metaphor processing region was determined by a conjunction null analysis (Nichols et al., 2005) reflecting common regions that responded more to multimodal metaphor videos than iconic videos, as well as, to monomodal metaphor videos than iconic videos. We also checked whether the greater activation for external than internal metonymy resolution (*ext* > *int*) had the slightest possible overlap with the metaphoric processing map by testing at a very liberal voxelwise-only threshold of $p < 0.05$, uncorrected.

Otherwise, results reported passed a voxel-wise statistical threshold

of $p < 0.001$ and were required to be part of a contiguous cluster of at least 34 significant voxels. This cluster volume threshold was calculated from 10,000 iterations of a Monte Carlo simulation using AFNI's 3dClustSim (<http://afni.nimh.nih.gov>) assuming an individual voxel type I error of $p < 0.001$ and the nearest neighbor clustering method (6 voxels) for a 5% probability of a cluster surviving the threshold by chance. All reported contrasts are one-sample t -tests that have also been masked by the constituent conditions above the low-level baseline, a visual crosshair, and thus reflect true activations larger than the low-level baseline. That is, for the contrast $A > B$ reported here, we masked with $A > \text{baseline}$ and $B > \text{baseline}$ to isolate activations above baseline (i.e. to exclude deactivations). Masks were thresholded at $p < 0.05$ and all resulting masked maps were thresholded as above. The conjunction null maps include voxels that are significant for each constituent contrast, thresholded at $p < 0.01$ and a minimum cluster extent threshold of 34 voxels. All coordinates are reported in MNI space, and images are displayed according to the neurological convention of left-side of image corresponding to anatomical left. Anatomical localization was based on the automated anatomical labeling atlas (AAL, Tzourio-Mazoyer et al., 2002).

2.6.3. ROI analysis

For a more detailed inspection of the differences between the two abstract cognitive processes of metonymy resolution and metaphor processing, we also calculated the percent signal change in atlas-defined ROIs, which were analogous to our functionally clusters of activation for metonymy resolution and supramodal metaphor processing. Percent signal change calculation was based on all significantly thresholded voxels below $p < 0.001$ and clusters with a minimum size of 34 voxels, within a given anatomical map-defined mask (detailed below) as the search volume. We used the *rfxplot* toolbox (Gläscher, 2009), which scales the percent signal change relative to the actual mean signal in a given ROI rather than the whole brain mean signal.

The whole-brain analysis results from comparing *ext > int* overlapped primarily with the more medial and ventral part of the (mirrored) right posterior TPJ (pTPJ) ROI defined by structural connectivity analysis in Mars et al. (2012). The cluster of activation we found for supramodal metaphor processing extended into the a/pMTG of the Harvard-Oxford atlas (Desikan et al., 2006), but was primarily located in the a/pSTG. Accordingly, we created masks for the pTPJ Mars et al. (2012), from the TPJ structural connectivity analysis from, and aSTG and pSTG (from the Harvard-Oxford cortical atlas: Desikan et al., 2006). We also created a left pars triangularis (IFGtr) mask from the Harvard-Oxford atlas, following the findings of this area in metonymy resolution in written text (Rapp et al., 2011). Both atlases were distributed with the FSL software package (Smith et al., 2004). To aid interpretation of the differences in brain activity for *ext* compared to *int* resolution, we correlated the percent signal change for *ext > int* in these four ROIs with the difference of the ex-Gaussian parameters estimated for each participant's RT distributions to *ext* and *int* items. We also tested correlations between percent correct concrete/abstract responses with percent signal change difference in these four ROIs. Pearson's product moment correlation was used for all correlation tests except with those involving percent correct concrete/abstract responses, which were analyzed with Spearman's rank correlation test.

2.6.4. Effective connectivity modeling

To test our hypothesis that metonymy resolution modulates semantic processing, we also created dynamic causal models (DCMs). We specifically hypothesized that the TPJ, which we found to be more active for *ext* than *int* processing, modulates activity implicated by metaphoric processing in the STG/MTG. This is plausible given previous studies which have reported TPJ's modulatory role, as reviewed in the Introduction (Section 1.4). These studies taken together with the results on metonymy processing in both language, apraxia and developmental studies also reviewed in the Introduction, set a precedent for

our effective connectivity analysis of whether, indeed, the TPJ, perhaps up-regulated in the more “complex” *ext* resolution, modulates abstract semantic networks.

For the DCMs, we extracted the signal in an 8-mm radius sphere centered on peak coordinates of three functionally defined volumes of interest (VOIs). Given that DCMs are based on model comparisons, there is no concern of double-dipping although two of the VOIs resulted from the same data. The three functional VOIs chosen were (1) the left TPJ region involved in greater activity for metonymy type resolution based on the results of this study (*ext > int* map, MNI peak coordinates $[-45, -52, 22]$), (2) the left STG/MTG for supramodal metaphor processing reported in a separate report (peak coordinates $[-57, -19, 1]$ from Joue et al., in press), and (3) the left pars triangularis (IFGtr) or more dorsal part of the IFG (dIFG) showing greater activity for metonymy resolution than for literal written sentences reported in Rapp et al. (2011) (peak coordinates $[-42, 24, 18]$). We only analyzed the effective connectivity between left-lateralized functional TPJ, IFG and STG/MTG to reduce complexity. Time-series data were extracted with the SPM8 eigenvariate toolbox. Each participant's principal eigenvariate of the BOLD time-series were calculated around individual local maxima within an 8-mm spherical radius of the global peaks of these VOIs and masked by an anatomical map from the Harvard-Oxford cortical atlas (Desikan et al., 2006). Time-series were mean-corrected.

Twelve dynamic causal models (two families of six models) were tested (Fig. 7). One family tested metaphor processing as bottom-up and metonymy resolution as top-down activity modulation. The other family tested metonymy resolution as bottom-up and metaphor processing as top-down activity modulation. Models within each family differed in intrinsic/anatomic connections and modulation inputs. Models between the two families had exactly the same structure and differed only in whether metaphor processing or metonymy resolution served as driving or modulating input.

As mentioned, because all communicative coverbal gestures reflect either internal or external metonymy, it is not possible to design conditions completely separating gesture metonymy resolution from multimodal metaphor processing. To compensate for this, all metaphor GLM regressors, collapsed over metonymy (i.e. *mono.ext*, *mono.int*, *multi.ext*, *multi.int*) were used to represent metaphoric processing, regardless of modality. GLM regressors *ico.int* and *ico.ext* were then used to represent metonymy with no metaphor processing. In models 1–6, “metaphoric processing” were the driving inputs to the left STG/MTG, and models differed by where metonymy processing modulated activity (explicitly shown in Fig. 7). Models 6–12 had analogous connections to models 1–6 except metonymy was the driving input and metaphoric processing modulated connections.

In two models in each family, all ROIs were connected intrinsically except between IFGtr and TPJ, following lack of reports of this connection in language-related fiber tracking (Gierhan, 2013) and DTI studies on resting state functional connectivity (Mars et al., 2012). These studies have reported, however, that the inferior longitudinal fascicle links the pSTG and parietal areas, and the uncinate fascicle and extreme capsule fascicle between IFG and pSTG (for a review, see Gierhan, 2013). Given the adjacency of the TPJ and IFG to long-range fiber tracts (Gierhan, 2013; Ramayya et al., 2010; Saur et al., 2008), such as the extreme capsule fascicle which appears to run between the IFG, STG and inferior parietal areas (Makris and Pandya, 2009), we also tested four models where all ROIs were intrinsically fully connected.

Models were estimated and compared in DCM10 of SPM8. Group-level parameters (Bayesian weighted connection strengths and posterior probabilities) were first calculated using random effects analysis to account for any cognitive processing differences between subjects. Bayesian Model Selection (BMS), based on the Bayes Factors of the negative variational free energy, was used to select the best fitting model (Penny, 2012).

Table 2

Peak coordinates comparing *int* vs. *ext* averaged across metaphoricity/non-metaphoricity and within multimodal metaphoric videos.

Contrast	Region	Side	Size	MNI coords [mm]			t-stat
				x	y	z	
<i>ext</i> > <i>int</i>	TPJ/pMTG/pSTG	L	115	−45	−52	22	4.76
<i>multi. ext</i> > <i>multi. int</i>	TPJ/pMTG/pSTG	L	158	−45	−61	7	5.72
	IFGtr/IFGop	R	56	45	8	28	4.72

A left TPJ was also found to survive threshold for *ico. ext* > *ico. int* but were deactivations and is thus not shown. All clusters of activity in this table have been consistently reported in semantic tasks in linguistic and action (as reviewed in the Introduction), as well as, the one prior linguistic metonymy study (Rapp et al., 2011). Clusters survived voxelwise threshold of $p < 0.001$ and a cluster size threshold corrected at $p < 0.05$, also subdivided according to metaphoricity; that is, external metonymy (*ext*) and internal metonymy (*int*) versus iconic gestures (*ico*), multimodal metaphoric expression (*multi*) and monomodal metaphoric gesture (*mono*). Peak activations are reported with corresponding hemispheric side (Side), and MNI coordinates (x y z). Cluster sizes are in number of voxels (Size). All activations reported have been masked by activations above the baseline and thus only reflect positive activations. Abbreviations: posterior (p), gyrus (G), pars opercularis (IFGop) of the inferior frontal gyrus (IFG), angular gyrus (AG), temporoparietal junction (TPJ), middle temporal gyrus (MTG), superior temporal gyrus (STG).

3. Results

The detailed results on multimodal metaphoric processing and grooming gestures are out of the scope of this paper. We report, instead, the direct differences between metonymy and metaphoric processing, and compare the metonymy types: internal (*int*; BPO) and external metonymy (*ext*; pantomimed).

3.1. Behavioral data

There was no difference in mean RTs to the “Was the topic of the video concrete” question after each video for *ext* ($M = 796.9$ ms, $SD = 498.7$ ms) and *int* ($M = 795.7$ ms, $SD = 482.1$ ms).

Metonymy type did not affect abstractness judgment nor interacted with video category (main effect of metonymy type: $F(1,31) = 2.46$, $\eta = 0.001$, $p = 0.13$, interaction $F(1.85,57.35) = 1.41$, $p = 0.25$). A more detailed report on further RT analyses can be found in the Suppl. Info.

3.2. fMRI data

A left TPJ cluster extending into the posterior MTG/STG cluster elicited significantly greater positive activations for *ext* than *int* gestures (Table 2, Fig. 4). This cluster is much more posterior to the MTG/STG activation for supramodal metaphoric processing (Fig. 5; see Methods, Section 2.6). When only considering iconic gestures, this left TPJ cluster survived voxel and cluster thresholds but did not involve only positive activations (did not survive masking *ico. ext* > *ico. int* by *ico. ext* > baseline and *ico. int* > baseline). When only considering multimodal metaphoric videos (*multi*), there was greater activation in the right IFGop/IFGtr for *ext* than *int* gestures (i.e. *multi. ext* > *multi. int*).

Differences in percent correct responses for *ext* and *int* only had very weak and unreliable correlations with activity in the left pSTG and left pTPJ (Pearson's $\rho = 0.15$, $p > 0.1$). Differences in σ estimates for *ext* and *int* were moderately correlated with differences in *ext* and *int* activity in all the ROIs (Pearson's $\rho = -0.22$ for left pTPJ, $\rho = -0.25$ for left aSTG, $p > 0.01$), but strongest in the left pSTG (Pearson's $\rho = -0.27$, $p < 0.01$).

The conjunction analysis between the *ext* > *int* contrast map and supramodal metaphoric processing map did not reveal any surviving clusters, not even at the liberal $p < 0.05$ uncorrected threshold.

According to Bayesian model selection on the DCMs, model (b) was the best fitting model (Fig. 8). Mean parameter estimates for this winning model are shown in Fig. 7. They indicate that pSTG (associated with metaphor processing) drives forward connections to the TPJ (associated with upregulated activity in metonymy resolution), which in turn receives backward inhibitory connections from dIFG. The dIFG also has strong backward inhibitory connections to pSTG, which is strengthened by the TPJ.

4. Discussion

Metonymy is an example of drawing inferences from symbolic representations either of real-world objects (first order representation) or with other symbols (higher order representation), where symbolic representations are by definition abstract. As explained in the Introduction, external metonymy resolution at its simplest requires second-order abstract inferencing: inference of the unrepresented object and of the bodily actions associated with the object. In contrast, internal metonymy resolution is at its simplest a first-order inference: inference of the part of the object represented.

As reviewed in the Introduction (Section 1.3), most studies finding

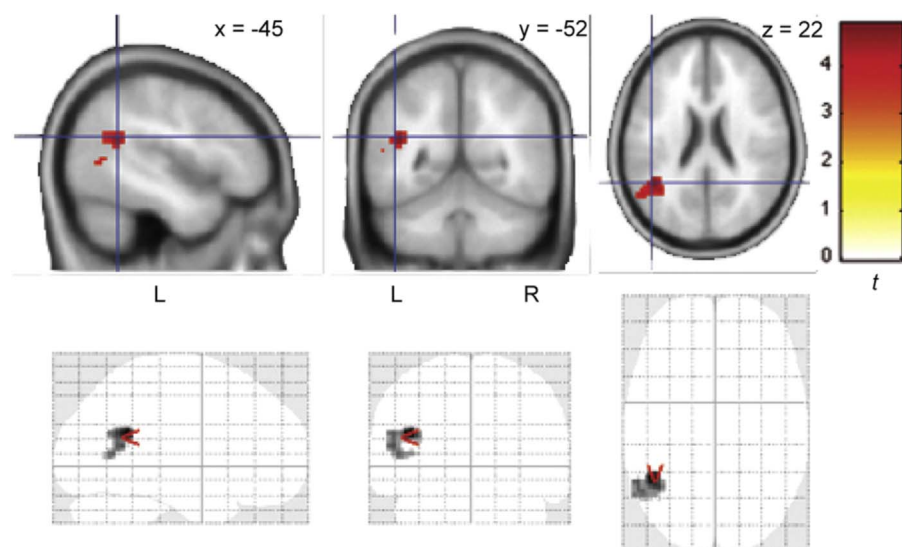


Fig. 4. External metonymy (*ext*) vs. internal metonymy (*int*). A TPJ/pMTG cluster was found to be more active for external metonymy/pantomimed gestures than internal metonymy/BPO gestures (*ext* > *int*; increasing positive activations shown from yellow to red). This suggests possibly greater processing demands for *ext* gestures compared to *int* gestures, such as greater attention required for semantic associations. Alternatively, it supports the role of the TPJ in establishing higher order or global inferences. Thresholds and abbreviations as in Table 2.

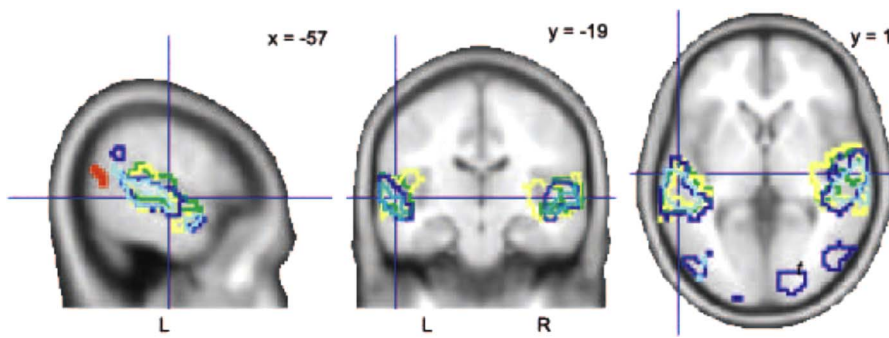


Fig. 5. Metonymy vs. Metaphor. Greater activation was found in left TPJ/pMTG for ext > int (red), which was more posterior than any of the clusters for metaphor processing, subdivided into metonymy type (see Suppl. Info Table S1). Light colors are for ext and darker colors for int: mono.ext > mono.int (light blue), mono.int > mono.int (royal blue), multi.ext > multi.int (light green), multi.int > multi.int (green). Crosshairs are located at the activation peak for the left MTG/STG found for supramodal metaphoric processing (MNI coordinates shown). Thresholds and abbreviations as in Table 2.

differences between external versus internal metonymy in gestures have been done in apraxia or in young children, perhaps as no differences are expected in healthy adults. Indeed, we found very little behavioral differences in our task (which we will elaborate on in the following discussion), but we found support for our theory of metonymy and inference complexity from the neuroimaging data. Specifically, we found greater activation in the left TPJ for external metonymy resolution compared to internal metonymy resolution (ext > int), which could arguably reflect greater processing required for the increased number of semantic attentional shifts and/or inference steps. However, it could also be argued that external metonymy is more commonly used in healthy adults, and greater activation could reflect greater sense of understanding of coverbal gestures reflecting external metonymy. However, the comparable variability in RTs to abstract/concrete responses in both metonymy types and perhaps greater variability for ext according to ex-Gaussian modeling, is inconsistent with this latter interpretation (see Suppl. Info Fig. S1). Furthermore, our neuroimaging data in context of findings reported by other studies, are consistent with our theory that metonymy is a mechanism to focus interpretation on relevant features, local and global, of what is communicated.

4.1. Metonymy as means to focus on what is represented

This study can be set in the context of whether pantomiming difficulties, as is seen in left brain damaged individuals and in developmental patterns, is also reflected in coverbal gestures processed by healthy adults. Previous studies in brain damage patients (McDonald et al., 1994) and in child development (Piaget, 1962; Bates et al., 1980; O'Reilly et al., 1997; Dick et al., 2005) have found some associations between pantomiming difficulties and the cognitive processes for symbolic representations (as discussed in Section 1.3). Although inconclusive results have also been reported (Goldenberg et al., 2003), there is general consensus that interpreting pantomiming actions requires context and semantic associations (Osiurak et al., 2012). Coverbal gestures lie at the interface of linguistic comprehension and action understanding, where spoken linguistic information determines whether the accompanying manual gestures employed external (ext) or internal metonymy (int). In investigating metonymy in coverbal gestures, we extended metonymy research beyond tool-based actions.

The left TPJ cluster, which we found more active for ext (pantomiming) than int (BPO) gestures, is posterior to the left lateralized MTG found by Rapp et al. (2011) for metonymy processing in written sentences. Indeed, it overlaps with a cluster common to numerous linguistic semantic studies (Horwitz et al., 1998). It overlaps with the anteroventral TPJ reported by Xu et al. (2005) for discourse compared to sentential processing, as well, as with the supramarginal gyrus (SMG)/TPJ/pSTG cluster reported by Chow et al. (2008) for predictive inferences during discourse compared to reading strings of pseudowords. It is located just inferior of the angular gyrus cluster found by Obleser et al. (2007), which was associated with successful understanding of speech that was acoustically degraded.

The TPJ has been reported to be implicated in increased inferencing

needed in a range of contexts including understanding the semantics in language, intentionality and perspective taking (Kim et al., 2012), reading weakly related discourse scenarios than highly related or unrelated sentences (Kuperberg et al., 2006), reading verbs with definite end states (Romagno et al., 2012), and attributing beliefs and other mental states (e.g. Zaitchik et al., 2010). In a meta-analysis of semantic studies, Noonan et al. (2013) has argued for its role in semantic executive control. This view of the TPJ as having a modulatory control of semantic retrieval, perhaps attention control, is an attempt to provide a unifying account of the wide range of cognitive functions attributed to the TPJ. Our conjunction analysis results suggest that the supramodal metaphor and ext > int maps unlikely share areas of activations. Although caution should be taken with interpreting these negative findings and comparisons, the very liberal thresholding in the conjunction analysis suggests that the lack of common clusters of activation is not due to lack of power. Moreover, we found greater correct abstract/concrete responses to correlate with lower activity in the left pSTG and lower exponential standard deviations in RTs (σ of ex-Gaussian models). Correct responses in our data are influenced by video category type (iconic, multimodal metaphoric expressions, monomodal metaphoric gestures) but not by metonymy type, thus linking correct concrete responses, video category type, and RT standard deviations with pSTG activation. This supports the interpretation of the role of the pSTG in semantic processing, distinct from the role of the TPJ.

Our effective connectivity results are consistent with the view that the TPJ is a hub. Like previous studies that have shown, for example, that TPJ activity is highly correlated with activity in the pSTG and IFGop in healthy readers, a connectivity which is disrupted in dyslexic readers (Horwitz et al., 1998). The cluster we found also corresponds to the pTPJ found in an ICA analysis to be connected with ipsilateral STS, contralateral pTPJ, MTG, and frontal regions (Igelström et al., 2015). We propose that the TPJ could possibly play role in orienting attention to clues needed for semantic processing, although it is debatable whether semantic attention can be separated from semantic processing (c.f. Ravizza et al., 2011 whose verbal short term memory study suggests that STM is more likely a semantic storage while TPJ reflects attentional processes). This cognitive process is fundamental to complex situations such as social contexts and language understanding, both of which require inferencing and hence establishing potentially many layered and hierarchical semantic associations. This proposal is along the lines of the Nexus Model of the TPJ (Carter and Huettel, 2013) (although note that Mars et al., 2012 found in a resting-state connectivity analysis that the aTPJ is connected to areas implicated in attentional selection processes whereas the pTPJ shows connectivity with areas reported for social cognition) and ties in theoretical linguistic accounts of the mediating role metonymy has on metaphor understanding (Mittelberg and Waugh, 2009, 2014; Mittelberg and Joue, 2017).

The likely role of the TPJ as a cognitive processing hub and the top-down role of dIFG are supported in our DCM analysis: although greater activity was found in the TPJ for ext, which we have argued to involve a higher order of semantic inferencing, our DCM analysis favored a model where metonymy resolution exerts top-down modulation of the

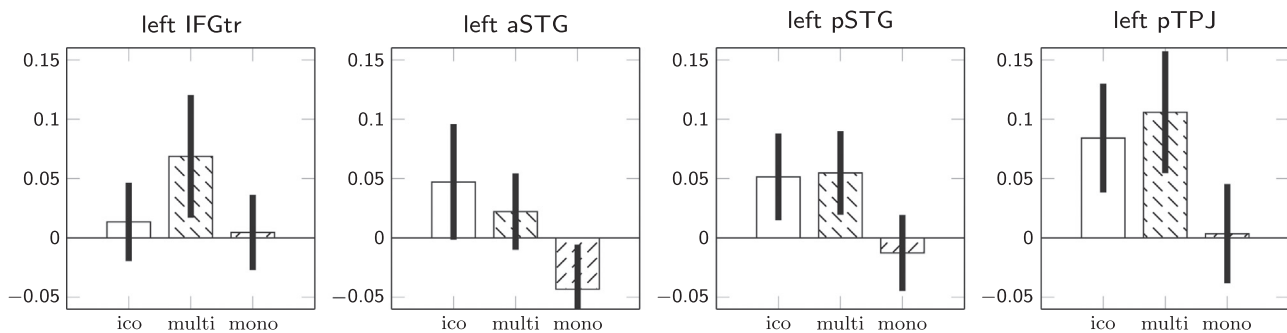


Fig. 6. Differences in percent signal change for ext > int in the anatomically defined ROIs left IFGtr, anterior and posterior STG and TPJ per video category (ico, multi, mono).

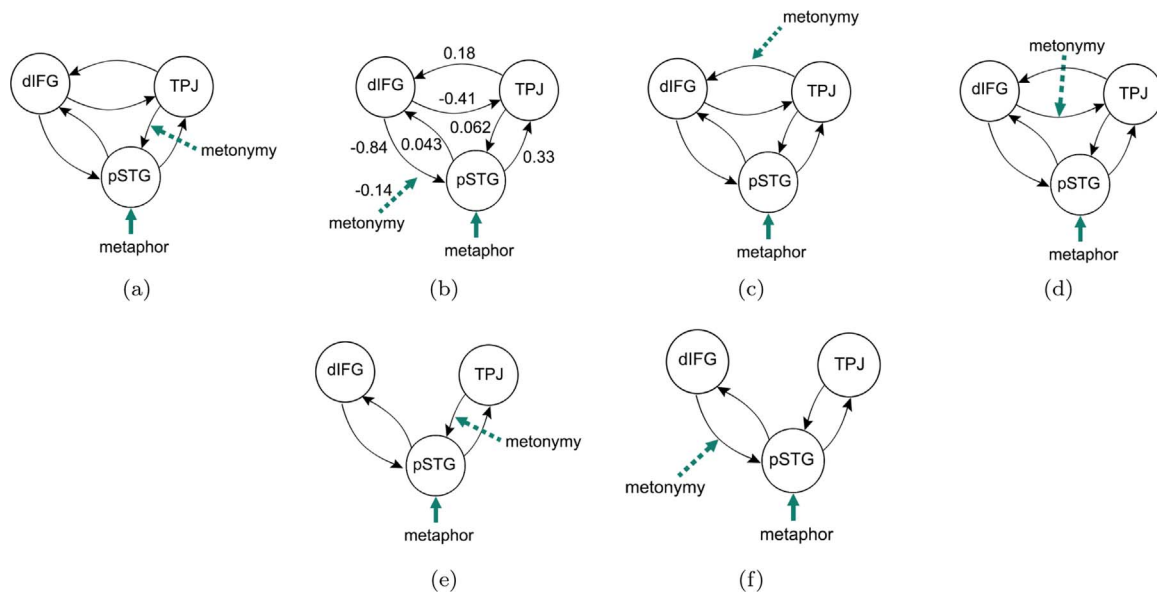


Fig. 7. Dynamic causal model space tested with metaphor processing as driving input and metonymy resolution modulating connections. In this model space, all models had metaphorical stimuli across modalities as driving inputs into pSTG. Models differed in intrinsic/anatomic connections (solid black arrows) and which of these connections were modulated by metonymy resolution (dotted green arrows). VOIs were functionally defined spheres around peak coordinates. Left inferior frontal gyrus (IFG): peak coordinates (−42, 24, 18) from Rapp et al. (2011). Left temporoparietal junction (TPJ): (−45, −52, 22) from ext > int cluster in this study. Left superior temporal gyrus (STG): (−57, −19, 1) from supramodal metaphoric processing contrast from Joue et al., (). Mean parameter estimates from Bayesian parameter averaging (BPA) are shown on the winning model, model (b). Although a total of twelve models were tested, only models 1–6 are shown here for brevity: models 7–12 have the same structure as 1–6, respectively, except metonymy resolution was specified as a driving input to pSTG, and metaphor processing modulated connections shown to be modulated by metonymy here.

dIFG on pSTG (Fig. 7). Semantic processing in pSTG drives activity in the TPJ, triggering activity in the dIFG. When semantic search is successful, dIFG inhibits activity in TPJ and constrains semantic search processing in pSTG. In this case, more semantic processing in STG is required than top-down attention inhibitory connections; semantic processing also needs to follow from ext or int attention selection. These reciprocal inhibitory and excitatory connections can explicate why activations for multi and mono do not show a clear trend for ext > int in PSCs in the three VOIs and the greater right IFGtr activation for multi.ext than multi.int but not for mono.ext or ico.ext (Fig. 6). Our DCM results also are consistent with the functional connectivity analysis in Willems et al. (2009) who found that left IFG modulated left pSTG activity. Greater right IFGtr activity is consistent with previous findings reporting its involvement in imagined tool use (Higuchi et al., 2007) and for interpreting object-directed actions, whether presented visually or in sentences describing the action (Baumgaertner et al., 2007).

It is in line with the role of the TPJ in multisensory bottom-up orienting of attention to stimulus features (Downar et al., 2000; Corbetta and Shulman, 2002; Cabeza et al., 2012; Shomstein, 2012), although these studies often report right TPJ whereas we found a left-lateralized cluster. Of note, the left TPJ has also been reported to be involved in attentional control but is seldom reported in detail (for a review, see

Geng and Vossel, 2013). Laterality has also been a debate in figurative processing. In a meta-analysis of action observation and imitation studies, however, Caspers et al. (2010) found that observation activated the pMTG bilaterally and only imitation tasks were left lateralized in the pMTG. This is in contrast to Ohgami et al. (2004) who found greater right SMG activation for executing right-handed BPO than pantomiming tool-use gestures, a result which is inconsistent with the developmental and aphasia studies which would predict increased demands for pantomiming gestures than BPO gestures and a left-lateralized network, both of which our results corroborate.

4.2. Metonymy as inferences on local or global semantic features

Our findings can also be situated in studies on global/local semantic processing. Simplistically, we can consider int resolution as requiring focus on more local details upon which semantic knowledge is accessed and inferences are drawn. In turn, ext can be considered as building on local details but directing attention to a more global picture integrating local details. Studies on global/local structure have moreover been attributed to the TPJ. Damage to the TPJ corresponded to attentional difficulties such as the inability to perceive global structures over local details in patients with TPJ lesions (Robertson et al., 1988; Bultitude et al., 2009) and in neuroimaging studies with healthy groups (e.g. Fink

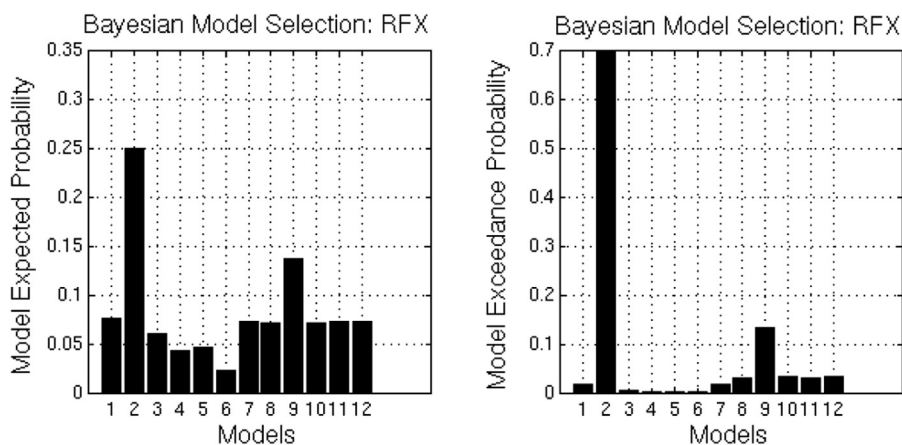


Fig. 8. Random-effects BMS results for the models considered in Fig. 7. Model 2 is clearly the best model.

et al., 1997a). Both lesion and neuroimaging studies have often used Navon letters to test attention to global/local structure, where repeats of a small letter of the alphabet are spatially arranged to form a larger letter. Again, these findings report a right-lateralized bias for processing global objects, but studies using non-linguistic stimuli have found bilateral TPJ activation (Fink et al., 1997b; Rennig et al., 2013). For example, given a full chess board and game-related assessment, greater bilateral TPJ activity is found in expert chess players compared to novice players who might focus more on individual chess pieces (Rennig et al., 2013). These results suggest that TPJ activity does not reflect semantic difficulty, in which case we could expect greater activity for the novice players. Instead, it suggests that the TPJ is involved in attention modulation for when prior semantic associations exist and could perhaps also reflect semantic representation updates. Using a letter-based Navon paradigm, bilateral TPJ activity correlated with length of sustained attention to either global or local level when participants did not know in advance which level to attend to (Fink et al., 1997a), a situation analogous to metonymy resolution in multimodal communication. This ties in with the linguistic view of coverbal gestures as offering components of a message (local features) that need to be synthesized (global feature), as discussed in the Introduction (see McNeill, 1992).

4.3. Limitations and future work

A limitation in our study is the lack of clear behavioral correspondence of inference complexity for *ext*. Future studies could focus on better defining the cognitive load in symbolic cognition complexity due to metonymy type, metaphor processing, and their interaction. The intrinsic undissociability of metonymy in coverbal gesture posed a difficulty in a meaningful baseline that isolates metonymy resolution in coverbal gestures. This undissociability also was the basis of the inseparability of gesture metonymy type from multimodal and monomodal metaphors. We were unable to directly compare metonymy type with metaphoric processing in multimodal communication, which is key to demonstrating that these are processes recruiting separate networks. Future studies could circumvent this issue by comparing gesture metonymy type with metaphoric processing in speech-only communication. However, comparison to a speech-only condition would not have permitted the separation of processes specific to audiovisual integration, such as possible superadditive effects, thus confounding our investigation of metonymy expressed in the coverbal gestures. Other gesture studies have often compared conditions against a baseline of rest, thus reporting a possible conflagration of processes involved in multimodal communication. The goal of this study was to investigate multimodal communication rather than comparing metaphor and metonymy in multimodal and unimodal communication. Nonetheless, separating modality-specific processes would certainly gain different

insights and should be further investigated.

Another limitation of this study is in our task. As mentioned, the judgment of abstractness/concreteness of concepts is known to be influenced by a complex interaction of numerous factors (Wiemer-Hastings and Xu, 2005; Ghio et al., 2013). Therefore, including this task possibly introduced cognitive processes that would not normally be present in language comprehension.

This is the first neuroimaging study to look at metonymy resolution, a form of abstract processing, in multimodal communication with coverbal gestures. Although our discussion remains speculative and our generalizations rest on backward inferences, we provide an integrative multidisciplinary account of symbol processing that is consistent with a view that the TPJ acts as a semantic hub, perhaps directing attention towards relevant information for inferencing and thus modulating the neural networks involved in accessing semantic knowledge. Metonymy in manual actions could very well be founded on the same abstract cognitive processes required for linguistic metonymy, where abstract cognition here refers to our ability to create representations of real-world objects, infer the objects from these symbols (first-order inference), and relate symbols with each other (second-order inference). Thus, the distinction between *ext* and *int* metonymy might appear subtle and indeed did not lead to much behavioral differences in young, healthy adults. However, our neuroimaging results show that metonymy resolution modulates metaphor understanding and suggest that metonymy resolution is fundamental and pervasive in daily, not just linguistic, cognition. It forms the basis upon which we create representations, or symbols, of our world.

Acknowledgments

The authors thank Ann Malzkorn and Christina Vedar for help with study materials; Alexandra Kleiman for aiding in initial data collection; and André Schüppen, of the Brain Imaging Facility of the Interdisciplinary Centre for Clinical Research within the Faculty of Medicine at the RWTH Aachen University, as well as, Thilo Kellermann for technical advice. **Funding**

This work was supported by the Excellence Initiative of the German Federal and State Governments and partially by the German Research Foundation (Deutsche Forschungsgemeinschaft DFG, IRTG 1328).

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.neuropsychologia.2017.12.033>.

References

- Badre, D., Poldrack, R.A., Paré-Blagoev, E.J., Insler, R.Z., Wagner, A.D., 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47, 907–918.
- Barcelona, A., 2009. Motivation of construction meaning and form: The roles of metonymy and inference. In: Panther, K.-U., Thornburg, L., Barcelona, A. (Eds.), *Metonymy and Metaphor in Grammar*. John Benjamins, Amsterdam/Philadelphia, pp. 363–401.
- Bates, E., Bretherton, I., Snyder, L., Shore, C., Volterra, V., 1980. Vocal and gestural symbols at 13 months. *Merrill-Palmer Q.* 26, 408–423.
- Baumgaertner, A., Buccino, G., Lange, R., McNamara, A., Binkofski, F., 2007. Polymodal conceptual processing of human biological actions in the left inferior frontal lobe. *Eur. J. Neurosci.* 25, 881–889.
- 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. *Cereb. Cortex* 19, 2767–2796.
- Binkofski, F., Buxbaum, L.J., 2013. Two action systems in the human brain. *Brain Lang.* 127, 222–229.
- Boyatzis, C., Watson, M., 1993. Preschool children's symbolic representation of objects through gestures. *Child Dev.* 64 (3), 729–735.
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A.M., Bölte, J., Böhl, A., 2011. The word frequency effect: a review of recent developments and implications for the choice of frequency estimates in German. *Exp. Psychol.* 58 (5), 412–424.
- Buckner, R.L., Koutstaal, W., Schacter, D.L., Rosen, B.R., 2000. Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. *Brain* 123 (3), 620–640.
- Bultitude, J.H., Rafal, R.D., List, A., 2009. Prism adaptation reverses the local processing bias in patients with right temporoparietal junction lesions. *Brain* 132, 1669–1677.
- Cabeza, R., Ciaramelli, E., Moscovitch, M., 2012. Cognitive contributions of the ventral parietal cortex: an integrative theoretical account. *Trends Cogn. Sci.* 16 (6), 338–352.
- Cabeza, R., Nyberg, L., 2000. Imaging cognition II: an empirical review of 275 PET and fMRI studies. *J. Cogn. Neurosci.* 12 (1), 1–47.
- Capek, C.M., Woll, B., MacSweeney, M., Waters, D., McGuire, P.K., David, A.S., Brammer, M.J., Campbell, R., 2010. Superior temporal activation as a function of linguistic knowledge: insights from deaf native signers who speechread. *Brain Lang.* 112, 129–134.
- Carter, R.M., Huettel, S.A., 2013. A nexus model of the temporal parietal junction. *Trends Cogn. Sci.* 17 (7), 328–336.
- Caspers, S., Zilles, K., Laird, A.R., Eickhoff, S.B., 2010. ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage* 50, 1148–1167.
- Chee, M.W., O'Craven, K.M., Bergida, R., Rosen, B.R., Savoy, R.L., 1999. Auditory and visual word processing studied with fMRI. *Hum. Brain Mapp.* 7 (1), 15–28.
- Chow, H.M., Kaup, B., Raabe, M., Greenlee, M.W., 2008. Evidence of fronto-temporal interactions for strategic inference processes during language comprehension. *NeuroImage* 40, 940–954.
- Clark, M.A., Merians, A.S., Kothari, A., Poizner, H., Macauley, B., Rothi, L.J.G., Heilman, K.M., 1994. Spatial planning deficits in limb apraxia. *Brain* 117, 1093–1106.
- Corbetta, M., Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L., 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* 3 (3), 292–297.
- Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215.
- Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., López, V., Núñez, R.E., 2009. Gesture and metaphor comprehension: electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain Cogn.* 70, 49–52.
- Covington, M.A., He, C., Brown, C., Naci, L., Brown, J., 2006. How complex is that sentence? A proposed revision of the Rosenberg and Abbeduto D-Level Scale. *Tech. Rep. CASPR Research Report 2006-01*, University of Georgia.
- De Renzi, E., Motti, F., Nichelli, P., 1980. Imitating gestures: a quantitative approach to ideomotor apraxia. *Arch. Neurol.* 37 (1), 6–10.
- Demal, U., 1999. *SKIDPIT-light Screeningbogen*. Universität Wien.
- Desikan, R.S., S'egonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968–980.
- Dick, A.S., Goldin-Meadow, S., Hasson, U., Skipper, J.I., Small, S.L., 2009. Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Hum. Brain Mapp.* 30 (11), 3509–3526.
- Dick, A.S., Mok, E.H., Beharelle, A.R., Goldin-Meadow, S., Small, S.L., 2014. Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. *Hum. Brain Mapp.* 35, 900–917. <http://dx.doi.org/10.1002/hbm.22222>.
- Dick, A.S., Overton, W.F., Kovacs, S.L., 2005. The development of symbolic coordination: representation of imagined objects, executive function, and theory of mind. *J. Cogn. Dev.* 6 (1), 133–161.
- Downar, J., Crawley, A.P., Mikulis, D.J., Davis, K.D., 2000. A multimodal cortical network for the detection of changes in the sensory environment. *Nat. Neurosci.* 3, 277–283.
- Elder, J.L., Pederson, D.R., 1978. Preschool children's use of objects in symbolic play. *Child Dev.* 49, 500–504.
- Fink, G.R., Halligan, P.W., Marshall, J.C., Frith, C.D., Frackowiak, R.S.J., Dolan, R.J., 1997a. Neural mechanisms involved in the processing of global and local aspects of hierarchically organized visual stimuli. *Brain* 120, 1779–1791.
- Fink, G.R., Marshall, J.C., Frith, P.W.H.C.D., Frackowiak, R.S.J., Dolan, R.J., 1997b. Hemispheric specialization for global and local processing: the effect of stimulus category. *Philos. Trans. R. Soc. Lond. Biol. Sci.* 264, 487–494.
- Frey, S.H., 2008. Tool use, communicative gesture and cerebral asymmetries in the modern human brain. *Philos. Trans. R. Soc. Lond. Biol. Sci.* 363, 1951–1957.
- Geng, J.J., Vossel, S., 2013. Re-evaluating the role of TPJ in attentional control: contextual updating? *Neurosci. Biobehav. Rev.* 37, 2608–2620.
- Ghio, M., Vaghi, M.M.S., Tettamanti, M., 2013. Fine-grained semantic categorization across the abstract and concrete domains. *PLoS One* 8 (6), e67090. <http://dx.doi.org/10.1371/journal.pone.0067090>.
- Gierhan, S.M., 2013. Connections for auditory language in the human brain. *Brain Lang.* 127, 205–221.
- Gläscher, J., 2009. Visualization of group inference data in functional neuroimaging. *Neuroinformatics* 7 (1), 73–82.
- Goldenberg, G., Hartmann, K., Schlott, I., 2003. Defective pantomime of object use in left brain damage: apraxia or asymbolia? *Neuropsychologia* 41, 1565–1573.
- Goodglass, H., Kaplan, E., 1963. Disturbance of gesture and pantomime in aphasia. *Brain* 86, 703–720.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Higuchi, S., Imamizu, H., Kawato, M., 2007. Cerebellar activity evoked by common tool-use execution and imagery tasks: an fMRI study. *Cortex* 43, 350–358.
- Hohle, R.H., 1965. Inferred components of reaction times as functions of foreperiod duration. *J. Exp. Psychol.* 69 (4), 382–386.
- Holle, H., Gunter, T.C., Rüschemeyer, S.-A., Hennenlotter, A., Iacoboni, M., 2008. Neural correlates of the processing of co-speech gestures. *NeuroImage* 39, 2010–2024.
- Holle, H., Obleser, J., Rueschemeyer, S.-A., Gunter, T.C., 2010. Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage* 49, 875–884.
- Horwitz, B., Rumsey, J.M., Donohue, B.C., 1998. Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc. Natl. Acad. Sci. USA* 95, 8939–8944.
- Igelström, K.M., Webb, T.W., Graziano, M.S., 2015. Neural processes in the human temporoparietal cortex separated by localized independent component analysis. *J. Neurosci.* 35, 9432–9445.
- Indovina, I., Macaluso, E., 2007. Dissociation of stimulus relevance and saliency factors during shifts of visuospatial attention. *Cereb. Cortex* 17, 1701–1711.
- Jakielski, K.J., 1998. *Motor Organization in the Acquisition of Consonant Clusters* (Ph.D. Thesis). University of Texas, Austin.
- Jakobson, R., Pomorska, K., 1983. *Dialogues*. MIT Press, Cambridge, MA.
- Joue, G., Boven, L., Willmes, K., Evola, V., Demeuscu, L. R., Hassemmer, J., Mittelberg, I., Mathiak, K., Schneider, F., Habel, U., in press. Metaphors are supramodal: the role of the bilateral lateral temporal regions in amodal abstract semantic processing.
- Kim, S., Yoon, M., Kim, W., Lee, S., Kang, E., 2012. Neural correlates of bridging inferences and coherence processing. *J. Psycholinguist. Res.* 41 (4), 311–321.
- Kimura, D., 1973. Manual activity during speaking. I. Right-handers. *Neuropsychologia* 11 (1), 45–50.
- Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., Konrad, K., Green, A., 2009. Neural interaction of speech and gesture: differential activations of metaphorical co-verbal gestures. *Neuropsychologia* 47, 169–179.
- Kuperberg, G.R., Lakshmanan, B.M., Caplan, D.N., Holcomb, P.J., 2006. Making sense of discourse: an fMRI study of causal inferencing across sentences. *NeuroImage* 33, 343–361.
- Lausberg, H., Zaidel, E., Cruz, R.F., Pfito, A., 2007. Speech-independent production of communicative gestures: evidence from patients with complete callosal disconnection. *Neuropsychologia* 45 (13), 3092–3104.
- Littlemore, J., 2015. *Metonymy*. Cambridge Studies in Cognitive Linguistics. Cambridge University Press.
- Makris, N., Pandya, D.N., 2009. The extreme capsule in humans and rethinking of the language circuitry. *Brain Struct. Funct.* 213, 343–358.
- Mars, R.B., Sallet, J., Schüffegen, U., Jabadi, S., Toni, I., Rushworth, M.F.S., 2012. Connectivity-based subdivisions of the human right Temporoparietal Junction Area: evidence for different areas participating in different cortical networks. *Cereb. Cortex* 22 (8), 1894–1903. <http://dx.doi.org/10.1093/cercor/bhr268>.
- Martin, A., Haxby, J.V., Lalonde, F.M., Wiggs, C.L., Ungerleider, L.G., 1995. Discrete cortical regions associated with knowledge of color and knowledge of action. *Science* 270 (5233), 102–105.
- McDonald, S., Tate, R.L., Rigby, J., 1994. Error types in ideomotor apraxia: a qualitative analysis. *Brain Cogn.* 25 (2), 250–270.
- McNeill, D., 1992. *Hand and Mind: What Gesture Reveals About Thought*. University of Chicago Press, Chicago.
- McNeill, D., 2005. *Gesture and Thought*. University of Chicago Press, Chicago.
- Mitchell, R.W., Clark, H., 2015. Experimenter's pantomimes influence children's use of body part as object and imaginary object pantomimes: a replication. *J. Cogn. Dev.* 16, 703–718.
- Mittelberg, I., Joue, G., 2017. Source actions ground metaphor via metonymy: toward a frame-based approach to gestural action in multimodal discourse. In: Hampe, B. (Ed.), *Metaphor: Embodied Cognition and Discourse*. Cambridge University Press, pp. 119–137.
- Mittelberg, I., Waugh, L.R., 2009. Metonymy first, metaphor second: A cognitive semiotic approach to multimodal figures of thought in co-speech gesture. In: Forceville, C. J., Urios-Aparisi, E. (Eds.), *Multimodal Metaphor*. Mouton de Gruyter, Berlin, New York, pp. 329–358.
- Mittelberg, I., Waugh, L. R., 2014. Gestures and metonymy. In: Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D., Bressen, J. (Eds.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*. Handbooks of Linguistics and Communication Science 38.2. De Gruyter Mouton, Berlin/Boston, pp. 1747–1766.
- Müller, C., Cienki, A., 2009. Words, gestures, and beyond: forms of multimodal metaphor

- in the use of spoken language. In: Forceville, C. J., Urios-Aparisi, E. (Eds.), *Multimodal Metaphor. Applications of Cognitive Linguistics*. Mouton de Gruyter, Berlin, pp. 297–328.
- Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.-B., 2005. Valid conjunction inference with the minimum statistic. *NeuroImage* 25, 653–660.
- Noonan, K.A., Jefferies, E., Visser, M., Ralph, M.A.L., 2013. Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *J. Cogn. Neurosci.* 25, 1824–1850.
- Oleser, J., Wise, R.J.S., Dresner, M.A., Scott, S.K., 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. *J. Neurosci.* 27 (9), 2283–2289.
- Ohgami, Y., Matsuo, K., Uchida, N., Nakai, T., 2004. An fMRI study of tool-use gestures: body part as object and pantomime. *NeuroReport* 15 (12), 1903–1906.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia* 9, 97–113.
- O'Reilly, A.W., 1995. Using representations: comprehension and production of actions with imagined objects. *Child Dev.* 66 (4), 999–1010.
- O'Reilly, A.W., Painter, K.M., Bornstein, M.H., 1997. Relations between language and symbolic gesture development in early childhood. *Cogn. Dev.* 12 (2), 185–197.
- Osiurak, F., Jarry, C., Baltenneck, N., Boudin, B., Le Gall, D., 2012. Make a gesture and I will tell you what you are miming. pantomime recognition in healthy subjects. *Cortex* 48 (5), 584–592.
- Panther, K.-U., Thornburg, L. L., (Eds.), 2003. *Metonymy and Pragmatic Inferencing*. John Benjamins, Amsterdam/Philadelphia.
- Pavani, F., Ládavas, E., Driver, J., 2003. Auditory and multisensory aspects of visuospatial neglect. *Trends Cogn. Sci.* 7 (9), 407–414.
- Pederson, D.R., Rook-Green, A., Elder, J.L., 1981. The role of action in the development of pretend play in young children. *Dev. Psychol.* 17 (6), 756–759.
- Penny, W., 2012. Comparing dynamic causal models using AIC, BIC and free energy. *NeuroImage* 59, 319–330.
- Piaget, J., 1962. *Play, Dreams and Imitation in Childhood*. Routledge & Kegan Paul, London.
- Piaget, J., Inhelder, B., 1971. *Mental Imagery in the Child*. Routledge, New York.
- Ramayya, A.G., Glasser, M.F., Rilling, J.K., 2010. A DTI investigation of neural substrates supporting tool use. *Cereb. Cortex* 20, 507–516.
- Rapp, A.M., Erb, M., Grodd, W., Bartels, M., Markert, K., 2011. Neural correlates of metonymy resolution. *Brain Lang.* 119 (3), 196–205.
- Rapp, A.M., Mutschler, D.E., Erb, M., 2012. Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. *NeuroImage* 63, 600–610.
- Ratcliff, R., Murdock Jr., B.B., 1976. Retrieval processes in recognition memory. *Psychol. Rev.* 83 (3), 190–214.
- Ravizza, S.M., Hazeltine, E., Ruiz, S., Zhu, D.C., 2011. Left TPJ activity in verbal working memory: implications for storage- and sensory-specific models of short term memory. *NeuroImage* 55 (4), 1836–1846.
- Raymer, A.M., Maher, L.M., Foundas, A.L., Heilman, K.M., Rothi, L.J.G., 1997. The significance of body part as tool errors in limb apraxia. *Brain Cogn.* 34, 287–292.
- Rennig, J., Bilalić, M., Huber, E., Karnath, H.-O., Himmelbach, M., 2013. The temporoparietal junction contributes to global gestalt perception—evidence from studies in chess experts. *Front Hum. Neurosci.* 7 (513). <http://dx.doi.org/10.3389/fnhum.2013.00513>.
- Robertson, L.C., Lamb, M.R., Knight, R.T., 1988. Effects of lesions of temporal-parietal junction on perceptual and attentional processing in humans. *J. Neurosci.* 8, 3757–3769.
- Romagnolo, D., Rota, G., Ricciardi, E., Pietrini, P., 2012. Where the brain appreciates the final state of an event: the neural correlates of telicity. *Brain Lang.* 123, 68–74.
- Rosenberg, S., Abbeduto, L., 1987. Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults. *Appl. Psycholinguist.* 8 (1), 19–32.
- Saur, D., Kreher, B.W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., Weiller, C., 2008. Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. USA* 105 (46), 18035–18040.
- Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26 (2), 100–105.
- Serences, J.T., Shomstein, S., Leber, A.B., Golay, X., Egeth, H.E., Yantis, S., 2005. Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychol. Sci.* 16 (2), 114–122.
- Shomstein, S., 2012. Cognitive functions of the posterior parietal cortex: top-down and bottom-up attentional control. *Front Integr. Neurosci.* 6 (38). <http://dx.doi.org/10.3389/fnint.2012.00038>.
- Shriberg, L.D., 1993. Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *J. Speech Lang. Hear. Res.* 36, 105–140.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., Stefano, N.D., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23, 208–219.
- Straube, B., Green, A., Bromberger, B., Kircher, T., 2011. The differentiation of iconic and metaphoric gestures: common and unique integration processes. *Hum. Brain Mapp.* 32, 520–533.
- Straube, B., Green, A., Weis, S., Chatterjee, A., Kircher, T., 2008. Memory effects of speech and gesture binding: cortical and hippocampal activation in relation to subsequent memory performance. *J. Cogn. Neurosci.* 21 (4), 821–836.
- Straube, B., He, Y., Steines, M., Gebhardt, H., Kircher, T., Sammer, G., Nagels, A., 2013. Supramodal neural processing of abstract information conveyed by speech and gesture. *Front. Behav. Neurosci.* 7. <http://dx.doi.org/10.3389/fnbeh.2013.00120>.
- Suddendorf, T., Fletcher-flinn, C., Johnston, L., 1999. Pantomime and theory of mind. *J. Genet. Psychol.: Res. Theory Hum. Dev.* 160 (1), 31–45.
- Turken, A.U., Dronkers, N.F., 2011. The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. *Front. Syst. Neurosci.* 5, 1–20. <http://dx.doi.org/10.3389/fnsys.2011.00001>.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15, 273–289.
- Van Overwalle, F., Mariën, P., 2016. Functional connectivity between the cerebrum and cerebellum in social cognition: a multi-study analysis. *NeuroImage* 124, 248–255.
- Vigneau, M., Beaucousin, V., Hervé, P., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., Tzourio-Mazoyer, N., 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *NeuroImage* 30, 1414–1432.
- Virtue, S., Haberman, J., Clancy, Z., Parrish, T., Beeman, M.J., 2006. Neural activity of inferences during story comprehension. *Brain Res.* 1084, 104–114.
- Vygotsky, L., 1976. Play and its role in the mental development of the child. *Sov. Psychol.* 5, 6–18.
- Wagner, A.D., Paré-Blagoev, E.J., Clark, J., Poldrack, R.A., 2001. Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31, 329–338.
- Wang, J., Conder, J.A., Blitzer, D.N., Shinkareva, S.V., 2010. Neural representation of abstract and concrete concepts: a meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* 31, 1459–1468.
- Werner, H., Kaplan, B., 1963. *Symbol Formation: An Organismic-Developmental Approach to Language and the Expression of Thought*. Wiley, New York.
- Wiemer-Hastings, K., Xu, X., 2005. Content differences for abstract and concrete concepts. *Cogn. Sci.* 29, 719–736.
- Willems, R.M., Özyürek, A., Hagoort, P., 2009. Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage* 47, 1992–2004.
- Wittchen, H., Zaudig, M., Fydrich, T., 1997. *SKID. Strukturiertes Klinisches Interview für DSM-IV. Achse I und II. Handanweisung*. Hogrefe, Göttingen.
- Xu, J., Kemeny, S., Park, G., Frattali, C., Braun, A., 2005. Language in context: emergent features of word, sentence, and narrative comprehension. *NeuroImage* 25, 1002–1015.
- Zaitchik, D., Walker, C., Miller, S., LaViolette, P., Peczeko, E., Dickerson, B.C., 2010. Mental state attribution and the temporoparietal junction: an fMRI study comparing belief, emotion, and perception. *Neuropsychologia* 48, 2528–2536.