

# Language in context: Characterizing the comprehension of referential expressions with MEG

Christian Brodbeck<sup>a,b,\*</sup>, Liina Pykkänen<sup>a,b,c</sup>

<sup>a</sup> Department of Psychology, New York University, New York, NY 10003, USA

<sup>b</sup> NYU Abu Dhabi Institute, Abu Dhabi, United Arab Emirates

<sup>c</sup> Department of Linguistics, New York University, New York, NY 10003, USA

## ARTICLE INFO

### Keywords:

MEG  
EEG  
Sentence comprehension  
Reference resolution  
Visual short-term memory

## ABSTRACT

A critical component of comprehending language in context is identifying the entities that individual linguistic expressions refer to. While previous research has shown that language comprehenders resolve reference quickly and incrementally, little is currently known about the neural basis of successful reference resolution. Using source localized MEG, we provide evidence across 3 experiments and 2 languages that successful reference resolution in simple visual displays is associated with increased activation in the medial parietal lobe. In each trial, participants saw a simple visual display containing three objects which constituted the referential domain. Target referential expressions were embedded in questions about the displays. By varying the displays, we manipulated referential status while keeping the linguistic expressions constant. Follow-up experiments addressed potential interactions of reference resolution with linguistic predictiveness and pragmatic plausibility. Notably, we replicated the effect in Arabic, a language that differs in a structurally informative way from English while keeping referential aspects parallel to our two English studies. Distributed minimum norm estimates of MEG data consistently indicated that reference resolution is associated with increased activity in the medial parietal lobe. With one exception, the timing of the onset of the medial parietal response fell into a mid-latency time-window at 350–500 ms after the onset of the resolving word. Through concurrent EEG recordings on a subset of subjects we also describe the EEG topography of the effect of reference resolution, which makes the result available for comparison with a larger existing literature. Our results extend previous reports that medial parietal lobe is involved in referential language processing, indicating that it is relevant for reference resolution to individual referents, and suggests avenues for future research.

## 1. Introduction

When language is used for communication, new information is not presented in a vacuum but is connected to information that is already known. Consequently, a fundamental device of language are expressions that invoke entities that are already known to the addressee (Lambrecht, 1994). For example, use of the definite noun phrase *the revolution* signals that the addressee should be able to identify which revolution in particular the speaker is talking about. The interpretation of referring expressions is thus a very fundamental process in language comprehension, allowing new meaning to be constructed on the base of known background information (see also van Deemter (2016)). And yet, the neural correlates of successful reference resolution remain largely uncharacterized. This is apparent from the absence of referential processing from recent models of the neural basis of sentence level language comprehension (e.g. Friederici, 2011; Hagoort and

Indefrey, 2014). Here we report on work that introduces a paradigm to investigate the neural basis of reference resolution. We provide evidence across three experiments and two languages that reference resolution in simple visual referential domains involves medial parietal cortex.

Previous research has shown that referential language processing is fast and incremental (Tanenhaus et al., 1995), takes into account a wide array of extra-linguistic sources of information (Chambers et al., 2002; Altmann and Kamide, 1999; Kamide et al., 2003; Altmann and Kamide, 2007) and can even affect syntactic parsing decisions (Tanenhaus et al., 1995; Spivey et al., 2002). Most of this evidence has come from studies using the so-called visual world paradigm, in which participants' eye movements are recorded while they follow instructions to perform various tasks with objects laid out in front of them (for reviews see Tanenhaus and BrownSchmidt (2008) and Huettig et al. (2011)). Studies that focused directly on reference

\* Correspondence to: Institute for Systems Research, University of Maryland, College Park, MD 20742, USA.  
E-mail address: [christianbrodbeck@nyu.edu](mailto:christianbrodbeck@nyu.edu) (C. Brodbeck).

resolution found that people typically move their eyes to the referent of an expression as soon as they have sufficient information to identify it (Eberhard et al., 1995; Sedivy et al., 1999). For example, when participants were asked to *Touch the starred yellow square* in a context with only one starred item, they moved their eyes to the relevant item on average 250 ms after the end of the word *starred*, significantly earlier than in contexts where more than one object had stars. However, eye movements constitute an indirect measure and do not necessarily reflect reference resolution to a unique item. When the first syllable of an utterance is compatible with two different continuations (e.g., *the beetle vs the beaker*), addressees distribute their fixations over both items (Alloppenna et al., 1998). This suggests that eye movements reflect attentional processes, guided by even partial word information, and before reference can be positively resolved.

Neural measures can provide complementary information to improve our understanding of the computational stages of reference resolution. In particular, electrophysiological measures can provide temporally precise signals, making it possible to measure the response to individual words in coherent language stimuli. A repeated finding from event related potential (ERP) studies is an N400 reduction to expressions resolving reference to an entity introduced in the previous sentence in a non-marked way (Burkhardt, 2006; Ledoux et al., 2007). Referential context can influence this response, suggesting that referential processing is occurring in the relevant time window between 250 and 500 ms (Ledoux et al., 2007). A later positive component has been associated with retrieval and updating when a prior referent is mentioned again (Van Petten et al., 1991). However, this observation is complicated by the fact that a similar response has also been associated with the introduction of a new discourse referent, when compared with reference to an existing one (Burkhardt, 2006, 2007). Yet another study that manipulated referential status through the article (*Kathy sat nervously in the cab on her way to the airport. A/ The cab...*) found no late component in either direction (Anderson and Holcomb, 2005). Together these results suggest that the late ERP component is sensitive to multiple factors and not yet interpretable as a direct measure of referential status per se. Another brain measure, which has achieved a higher degree of functional specificity, is a component related to referential ambiguity. A group of EEG studies have found that referentially ambiguous expressions evoke a sustained frontal negative-going event-related potential when compared to unambiguous controls (reviewed by Nieuwland and Van Berkum (2008)). This includes determiner-noun phrases like *the girl* after a story context that introduced two girls vs one girl (Van Berkum et al., 1999, 2003; Nieuwland et al., 2007a; Boudewyn et al., 2015) as well as pronouns matching two vs one previously introduced characters (e.g., *he* in *Ronald told Frank that he...* can refer to *Ronald* as well as *Frank*, whereas *he* in *Ronald told Emily that he...* can only refer to *Ronald*) (Nieuwland and Van Berkum, 2006).

An fMRI study localized the response to referentially ambiguous pronouns to multiple prefrontal and parietal cortices, while unambiguous pronouns were associated with higher activation in the inferior frontal gyrus bilaterally (Nieuwland et al., 2007b). While referential ambiguity involves a direct manipulation of reference resolution, other cognitive processes might be involved. The failure to find a referent for a pronoun could be associated with an activity increase in the basic referential search processes (searching harder) or a decrease (giving up), as well as activity in other regions recruited to deal with the ambiguity (Nieuwland and Van Berkum, 2008). A subsequent study that focused on the decision making process associated with assigning pronoun referents based on different cues such as gender and verb bias confirmed that this component of the task involves a broad network of frontal, parietal and temporal regions (McMillan et al., 2012). While these results thus indicate a broad set of regions that could be relevant for reference resolution, they might also include regions involved in higher order decision making processes.

Our investigation had the goal of testing for activity in these regions

that is immediately associated with successful reference resolution when higher order decision making requirements are minimal. A challenge for studying successful reference resolution is finding a suitable control condition that does not involve creation of new discourse referents while at the same time not introducing another contrast like referential failure or referential ambiguity. To this end we adapted the visual world paradigm that has been used for studying incremental reference resolution (Eberhard et al., 1995, see above) for MEG/EEG in three experiments, each with the same basic trial structure (see Fig. 3 for Experiment 1). In each trial, participants first saw a simple visual world display and then read a question about that display, presented word by word. Participants were required to answer each question with a yes/no button press, encouraging them to process the questions naturally with the goal of comprehending them and without drawing undue attention to the referential aspect of the task. We analyzed the neural response to simple referential expressions like *the blue heart*, comparing the same expressions in different referential contexts. In the example in Fig. 3 the adjective *blue* could resolve reference in a context with one blue item, but not in a context with two blue items. This contrast isolates reference resolution with minimal overt ambiguity. Based on EEG data collected concurrently with the MEG data reported here for Experiment 2, we previously reported that evoked potentials to reference resolving words reflect the location of the referent on the display around 333 ms after adjective onset, confirming that the manipulation is effective in engaging referential processing at the adjective (Brodbeck et al., 2015). MEG allowed us to measure brain activity associated with the processing of individual words with high temporal precision. This allowed us to track neural activity occurring during the processing of specific words in referential expressions, in contrast to fMRI which does not allow attributing activity temporally to individual words in a sentence.

While the fMRI study on referential ambiguity (Nieuwland et al., 2007b, see above) provides us with a broad set of regions of interest, there are also other relevant prior results. First, another fMRI study found that discourses involving two conjoined referents compared to discourses with one or two singular referents engaged medial and superior/lateral parietal regions (Boiteau et al., 2014). This result suggests that parietal cortex is involved in the representation of discourse referents because it is sensitive to the manner in which multiple discourse entities are introduced. Second, reference resolution is a fundamental component of processing coherent language, because coherence critically depends on repeated reference to the same entities. A meta-analysis found that the processing of coherent language is associated with increased activity in medial parietal, medial frontal and bilateral temporal areas (Ferstl et al., 2008). Together, these results highlight the parietal lobe as possibly relevant for referential language processing, since it is the only region that was reliably affected by all three contrasts. Almor et al. (2007) suggested that parietal involvement in referential processing could reflect recruitment of circuits originally devoted to perceptual organization, tracking multiple objects in space, for keeping track of multiple discourse referents. This connection is particularly relevant for our design, which used visuo-spatial referential domains, and predicts that reference resolution should be associated with parietal activity.

## 2. Experiment 1

Experiment 1 demonstrated the viability of the paradigm and established the main result. However, in order to keep the paradigm and the task simple, compromises were made which lead to some potential confounds. Those were addressed in Experiments 2 and 3. Each visual world display was composed of three colored shapes, providing the context for simple adjective-noun referential expressions such as *the blue heart*. The primary target stimulus was the color adjective, the first word that differed between trials in whether it resolved reference or not: following a display with only a single blue

shape, *blue* provided enough information to identify the referent of the first noun phrase (“adjective resolving”). Following a display with two blue shapes, *blue* was compatible with both blue items, delaying reference resolution until the subsequent noun *heart* (“noun resolving”). Prior to the color adjective, the trials of the different conditions were indistinguishable at all levels; what differed between conditions was only the relationship between the visual world display and the referential expression. In line with the visual world studies described above, we expected participants to resolve reference upon perceiving the word that provided enough information for this, i.e., the adjective in the adjective-resolving condition, and the noun in the noun-resolving condition.

## 2.1. Methods

### 2.1.1. Participants

The experiment was conducted at the Washington Square (New York City) campus of New York University. We collected data from 21 right handed native speakers of English (12 female) with no known neurological abnormalities (mean age 24.7, range 18–45 years). The protocol was approved by the Institutional Review Board, and subjects provided written consent before beginning the experiment.

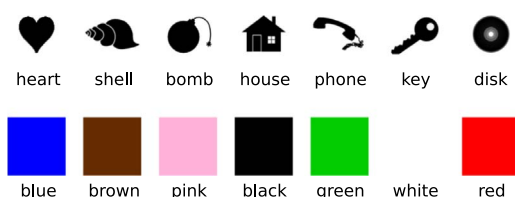
### 2.1.2. Design and stimuli

To construct the visual world displays, we identified 7 common, monosyllabic color adjectives, and 7 nouns matched to the adjectives for frequency and the number of characters (see Fig. 1). All displays contained three objects in a horizontal array and were constructed using two colors and two shapes (see Fig. 3). In each display, one of the three objects had a unique color and another object had a unique shape. The object with the unique color was always located on the outside of the display (either left or right).

Referential expressions were embedded in the beginning of questions conforming to the two schemas in (1) and (2), as illustrated in examples (3) and (4):

- (1) Was the ⟨color⟩ ⟨shape⟩ beside a ⟨shape⟩?
- (2) Was the ⟨color⟩ ⟨shape⟩ opposite a ⟨shape⟩?
- (3) Was the green key beside a bomb?
- (4) Was the green key opposite a bomb?

The relations, *beside* and *opposite*, were explained to participants before the experiment (*opposite* referring to the opposition of the two outside items). In half the questions, the color adjective identified a single item, whereas in the other half the display contained two items of the color, and the shape described by the noun allowed identification of the referent. The item(s) picked out by the color adjective could be located either on the left (and middle), or on the right (and middle). In contrast to Experiments 2 and 3, this allowed us to reduce the number of levels on the factor location from 3 (left, middle and right) to two (left and right), which allowed us to collect more trials per condition while keeping the overall experiment duration low, which was a concern after pilot subjects indicated that the task was tiring. Incidentally, this simplification also allowed us to make the factor location meaningful in both the resolving and the non-resolving condition, i.e., even when the adjective did not resolve reference its



**Fig. 1.** Experiment 1 stimuli: The 7 colors and the 7 shapes used in Experiment 1, labeled with the words used in the questions.

meaning singled out one side of the screen (left or right) over the other. The subsequent noun resolved reference to one of the two items singled out by the adjective with equal probability. This ensured that participants could not predict the referent of the noun after hearing a non-resolving adjective. Together with the fact that questions were about relationships to another item on the display, this assured that participants had to pay attention to the whole display rather than being able to strategically attend to only some information.

For the analysis of adjectives, this resulted in a 2 (reference resolution: resolving or non-resolving)  $\times$  2 (referent location: left or right) design. In each cell of this 2 $\times$ 2 design targets were constructed by permuting the 7 colors with the 7 shapes, resulting in 49 trials per condition and 196 trials in total. Our analysis thus compared linguistic expressions that were identical between conditions, with the conditions differing only in the relationship of the expression to the prior visual world context. The remaining aspects of the designs were pseudo-randomized but balanced between conditions, and the order of trials was randomized for each participant. On trials in which the noun resolved reference, the referent could also be located in the middle of the display (late referents left: 33; middle: 28; right: 37).

Fig. 3 illustrates the general trial structure. Each trial started with a fixation cross (presented for 600 ms), which was followed by the visual world display for 300 ms. The display was presented for a short duration in order to minimize the use of strategies, such as covert naming (Zelinsky and Murphy, 2000), and consistent with studies of visual short-term memory (Luck and Vogel, 1997). After an inter-stimulus interval (ISI) of 300 ms, a question followed in serial visual presentation. The words *was*, *the*, and *a* were presented for a duration of 200 ms and with a subsequent ISI of 200 ms; all other words were shown with a duration and ISI of 300 ms. The last word of the question together with the question mark stayed on the screen until the participant made a yes/no response by pressing one of two buttons. After the participants made a response, they received feedback (“correct” or “incorrect”) and the screen remained blank for an inter-trial interval randomly selected between 300 and 500 ms, followed by the next trial.

### 2.1.3. Procedure

Before the MEG recording, participants' head shape was digitized using an optical scanner (Polhemus FastSCAN, [www.polhemus.com](http://www.polhemus.com)). Five fiducial points were marked in the digital scans and coils were attached to those points on participants' face. The spatial position of those coils with respect to the MEG sensors was recorded before and after each MEG recording session, and were later used to determine the position of the participant's head with respect to the MEG sensors. The head shape scans were then used to position a virtual head and brain model with respect to the sensor position, in order to constrain the source localization of the MEG data.

Participants were given instructions on the reference task before entering the MEG acquisition chamber. They were allowed to perform as many practice trials as they wished until they felt comfortable doing the task (practice trials used stimuli that did not occur in the main experiment). Inside the magnetically shielded MEG acquisition chamber, participants lay in a supine position and stimuli were projected onto a screen at comfortable viewing distance above their eyes. The experimenters communicated with participants via a microphone and headphones. Participants were instructed to blink as little as possible during the presentation of the stimuli. They were told that if they needed a break they could withhold their yes/no response at the end of a trial until they felt comfortable to continue. In regular intervals throughout the experiment they were informed of the progress in the experiment with a text display and had the opportunity to take a short, self-terminated break. Stimuli were presented with MATLAB using psychtoolbox ([psychtoolbox.org](http://psychtoolbox.org)) and ptbwrapper ([code.google.com/p/ptbwrapper](http://code.google.com/p/ptbwrapper)). On average, the task lasted approximately 23 min from first to last trial.

### 2.1.4. MEG data acquisition and analysis

MEG data were collected using a 157-channel whole-head axial gradiometer system (Kanazawa Institute of Technology, Nonoichi, Japan) sampling at 1000 Hz with a low-pass filter at 200 Hz and a notch filter at 60 Hz. No high pass filter was applied during recording, however the electronics of the MEG sensors aiming to automatically offset the DC magnetic field lead to an attenuation of low frequencies, effectively equivalent with a 0.16 Hz high pass filter. Environmental noise was removed from the raw data by regressing signals recorded from three orthogonally oriented magnetometers approximately 20 cm away from the recording array against the recorded data using the continuously adjusted least squares method (Adachi et al., 2001).

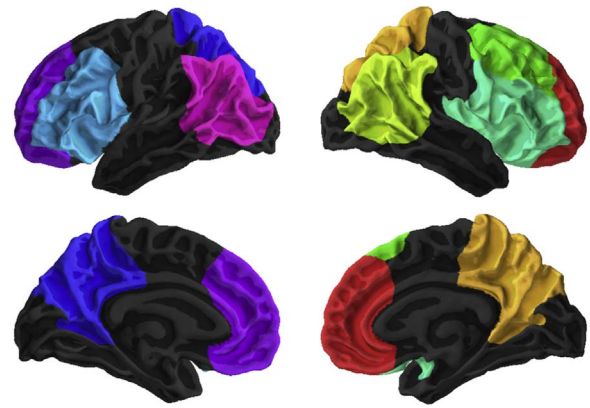
MEG data pre-processing, source reconstruction and statistical analysis were performed using mne-python (Gramfort et al., 2013, 2014) and Eelbrain ([pythonhosted.org/eelbrain](http://pythonhosted.org/eelbrain)).

**2.1.4.1. Preprocessing.** Data were band pass filtered between 1 and 40 Hz. The high pass filter was necessitated by the presence of high amplitude low frequency noise in the signal. Data were epoched in a time window from –100 to 600 ms relative to the onset of the adjective, and separately for the noun. For each subject, bad channels and epochs containing artifacts were excluded from further analysis. Epochs were downsampled to 200 Hz, averaged by subject and condition and baseline corrected with the 100 ms pre-stimulus interval.

**2.1.4.2. Source reconstruction.** Each participant's digitized head shape model was coregistered to the MEG sensor positions using the 5 fiducial points (see Section 2.1.3 above). A structural magnetic resonance image (MRI) processed with FreeSurfer software (<http://surfer.nmr.mgh.harvard.edu>) was then co-registered to the head shape. For 6 participants, structural MRIs were available; for the remaining participants an average brain provided by FreeSurfer (“fsaverage”) was substituted. Source spaces were defined on the white matter surface reconstructed with FreeSurfer using the topology of a recursively subdivided icosahedron (“ico-4”), resulting in 2562 sources per hemisphere.

At each source, free orientation dSPM noise-normalized minimum norm current estimates were computed (Hämäläinen and Ilmoniemi, 1994; Dale et al., 2000). Baseline noise covariance was estimated in the 100 ms baseline period of the adjective, and scaled with the default regularization factor (SNR=3). DSPM noise normalization (see Dale et al. (2000)) was used to improve dipole localization error, especially at deep sources compared to conventional minimum norm estimates (see Hauk et al. (2011)). For statistical analysis, data from all subjects were morphed to the fsaverage brain with a non-linear spherical transformation taking into account individual sulcal-gyral patterns (Fischl et al., 1999).

**2.1.4.3. Statistical analysis.** To evaluate the effects of our stimulus manipulation on the localized MEG activity, we employed spatio-temporal cluster tests within a broad search area defined on the basis of the fMRI study of referential ambiguity discussed in the Introduction (Nieuwland et al., 2007b). For each of the brain regions whose response differentiated between referentially ambiguous and unambiguous pronouns we used the peak coordinates provided by Nieuwland and colleagues. Peaks were located in medial parietal, lateral parietal, medial frontal and lateral frontal regions in each hemisphere and a right superior frontal region. For each peak, we selected the closest vertex on the fsaverage white matter surface as seed. Around those seed vertices, the search area was then grown on the white matter surface up to a geodesic distance of 50 mm. We excluded from the final search area midline sources lying in the corpus callosum and subcortical structures according to the PALS-B12 atlas



**Fig. 2.** Regions for the source space analysis: Regions in which the spatio-temporal cluster tests were performed, illustrated on the white matter surface of the fsaverage brain.

(Van Essen, 2005). The final search area is shown in Fig. 2, colored according to the location of the original seed vertices.

Within those regions we analyzed current estimates using spatio-temporal cluster based permutation tests to maintain strong control over type I error (Nichols and Holmes, 2002; Maris and Oostenveld, 2007). For each relevant statistical comparison, we computed maps of  $F$ - or  $t$ -values over space and time. We thresholded these maps at a  $F$ - or  $t$ -value equivalent to an uncorrected  $p=0.05$ . We then formed clusters of suprathreshold values adjacent in space and time. Adjacency in space was determined based on neighborhood of sources along the cortical surface. In addition, because surface connections in the FreeSurfer anatomical model do not cross the midline, each source vertex along the midline was connected to the closest vertex in the opposite hemisphere if that vertex was less than 15 mm away. This was done to better capture medial clusters of activation, which would be expected to involve sources in both hemispheres due to the spatial dispersion of MEG source localization (see Hauk et al. (2011)). As critical cluster statistic we used the exceedance mass, i.e. the sum of all  $t$ - or  $F$ -values in the cluster, which has the advantage of being sensitive to extension as well as magnitude of the effect (Nichols and Holmes, 2002). We computed distributions for this statistic by collecting the maximal statistic in 10,000 random permutations of the original data, shuffling the condition labels of the condition averages within subjects. In each permutation, one maximum cluster statistics was collected across the whole search region to correct the resulting  $p$ -values for multiple comparison across the whole search region. For each cluster in the original comparison we thus calculate a  $p$ -value as the proportion of the permutation distribution above the cluster's own exceedance mass.

We did not employ a step-down procedure after finding a significant cluster (see Nichols and Holmes (2002)). Consequently, when a comparison yielded more than one clusters to be compared against the same permutation distribution,  $p$ -values for all but the largest cluster are very conservative; for this reason we also mention in the results secondary clusters that reached the level of a trend ( $p \leq .1$ ).

Our time window of analysis extended from 200–500 ms. We chose the beginning of the time window to be able to capture relatively early effects, and the end was based on the observation in eye tracking studies that people fixated on the referent object approximately 550 ms after the onset of a disambiguating adjective (Eberhard et al., 1995). Allowing for necessary saccade planning (see Hutton (2008)) this suggests that identification of the referent should have occurred by 500 ms.

In order to maximize the power of our analysis we did not introduce an exclusion criterion based on performance. This was also justified because incorrect answers could have many reasons, only one of which was failure to resolve the question-initial referential expression.

Indeed, identification of the referent was one of the earlier cognitive operations required in answering the questions. The probably more difficult aspect of the task was to perform the mental comparison of the relational statement, indicating which shape was *next to* or *opposite* the referent, with the display. Many incorrect answers were probably due to failure in this task despite successful identification of the referent, and the trials would thus still be suited for analyzing the neural correlate of the resolution of the initial referential expression.

**2.1.4.4. Whole brain plots.** To visualize whether any other brain region might have been involved in reference resolution at a lower statistical threshold we also created whole brain plots of the relevant comparison. We computed a spatio-temporal map of *t*-values comparing the reference resolving to non-resolving adjectives, equivalent to the main effect of reference in the ANOVA. We thresholded the *t*-maps at a level equivalent to  $p=.05$  (uncorrected), only retaining clusters that spanned at least 100 sources and 50 ms. We then collected the extrema of the resulting *t*-maps in 100 ms bins for visualization.

## 2.2. Results and discussion

### 2.2.1. Behavioral results

On average, participants answered 85.9% of the questions correctly. Individual performance ranged from 67.9% to 97.4% correct. The number of mistakes did not differ significantly between trials in which the adjective or the noun resolved reference ( $t_{20}=0.95$ ,  $p=.352$ ), and referent location did not significantly influence performance in either condition ( $p_s \geq .151$ ).

### 2.2.2. MEG results

The analysis with a 2 (reference resolution) by 2 (location: left vs right) ANOVA yielded two significant clusters with a main effect of reference in the medial parietal lobe region, characterized by an increase in activity when reference could be resolved (255–325 ms,  $p=.036$  and 435–500 ms,  $p=.046$ ). Fig. 3 (top panel) illustrates the spatial extent of the clusters along with the time course of activation in the regions defined by their extent. Both clusters included sources in both hemispheres. While it is possible that the effect was caused by bilateral medial parietal activation, the medial parietal sources of the two hemispheres lie in close proximity, often less than 15 mm apart; in view of the spatial dispersion of MEG source estimates it would be premature to draw any conclusions about the lateralization of the effect, and a true source in either hemisphere could result in an effect in both hemispheres (see Hauk et al. (2011)). Fig. 4 illustrates differences between reference resolving and non-resolving adjectives across the whole brain.

The design of Experiment 1 was symmetric in that whenever the adjective did not resolve reference, the subsequent noun did. Consequently, we could also compare reference-resolving nouns with non-resolving nouns (the latter being cases where reference had already been resolved by the adjective). Including the location factor in the design would have led to an unbalanced model with an empty cell, because in trials with resolving adjectives, references had always been resolved towards the left or right item, whereas nouns also resolved reference to the middle with equal probability as to either side. For this reason we collapsed across location condition for the purpose of analyzing the nouns. Based on the location and timing of the two clusters in the response to the adjective we performed a one-tailed *t*-test in the medial parietal region in the time window from 250–500 ms (Fig. 3, bottom panel). This analysis revealed one significant cluster (340–445 ms,  $p=.042$ ).

In sum, the combination of our adjective and noun analyses yielded a consistent effect pattern in medial parietal cortex, with an increase in

activity whenever reference could be resolved.

Before these results can be attributed to reference resolution, however, it is necessary to disentangle potentially covarying processes. Specifically, reference resolving adjectives also enabled immediate pre-activation of the subsequent noun: matching the word *blue* to a picture of a blue heart forms a strong prediction that the noun will be *heart*. If such predictive processes take place on the adjective, one would expect less corresponding neural activity on the subsequent noun, resulting in exactly the pattern that we observe in our data: more activity on the resolving and predicting adjectives followed by less activity on the non-resolving, pre-accessed nouns. Neurobiologically, this hypothesis is at least plausible, given that medial parietal cortex has been argued to be part of a “semantic network” (Binder et al., 2009). To address this issue, we conducted two follow-up experiments to assess the extent to which the effects obtained in Experiment 1 might be explainable in terms of lexical prediction of the post-adjectival noun. In Experiment 2 this was achieved by a full crossing of the factors reference resolution and prediction and in Experiment 3, we conducted an Arabic version of Experiment 1, where, due to a different word order, reference resolution did not covary with lexical prediction.

## 3. Experiment 2

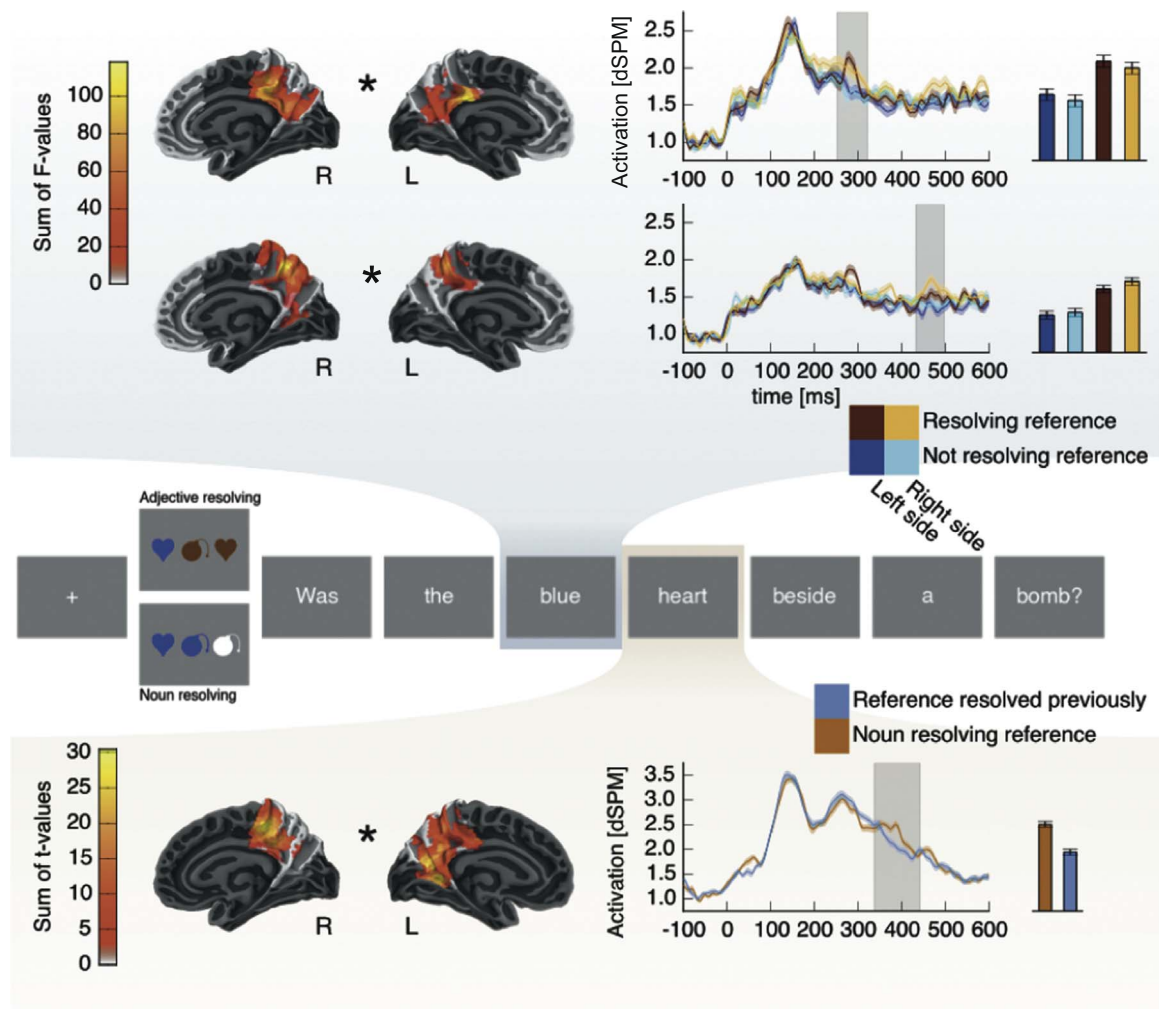
Experiment 2 was designed to answer the question: Is medial parietal lobe involved in processing the referential or predictive properties of reference-resolving adjectives? In order to disentangle reference resolution from prediction, we modified our design from Experiment 1 to include situations in which reference is resolved towards an occluded object whose shape cannot be determined, thus preventing lexical pre-activation. To achieve this, we added to our visual world displays large squares with an aperture. This square was normally behind the objects, but on certain displays it was raised in front of an object (see Fig. 6 on page 23). In those cases, the object's color was visible but not its shape. This allowed for a condition in which the color adjective identified the referent, but did not allow predicting the noun (reference without prediction). In order to construct the opposite condition, in which the adjective predicted the noun but did not identify the referent, we included displays containing two objects with identical color and shape which only differed in the visual pattern (with or without stripes). In the relevant trials, the color adjective did not resolve reference, because there were two objects with the same color, but it allowed predicting the noun, because both objects had the same shape (prediction without reference).

For a subset of participants in Experiment 2 we recorded EEG data concurrently with the MEG recordings. These data are used here only to test whether the main effect associated with reference resolution could also be measured with EEG. An orthogonal analysis of these data has been previously reported in the context of a different research question (Brodbeck et al., 2015).

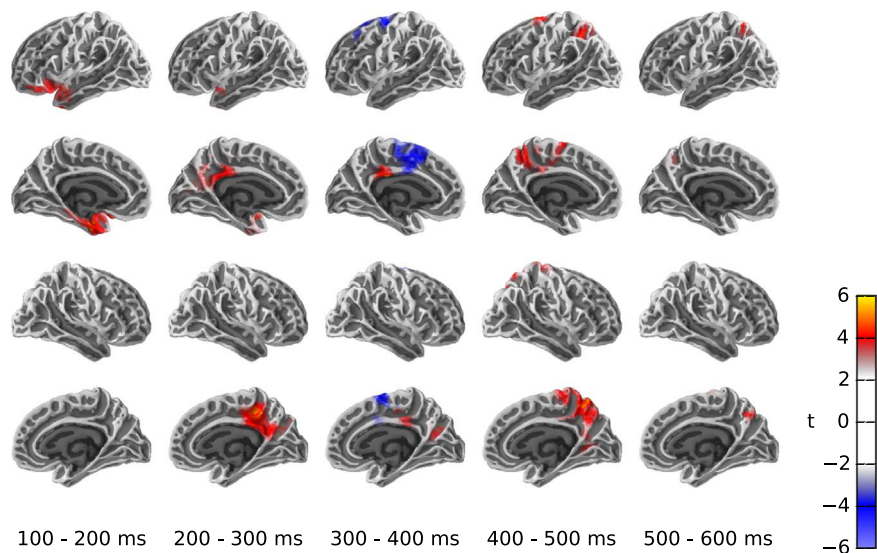
### 3.1. Methods

#### 3.1.1. Participants

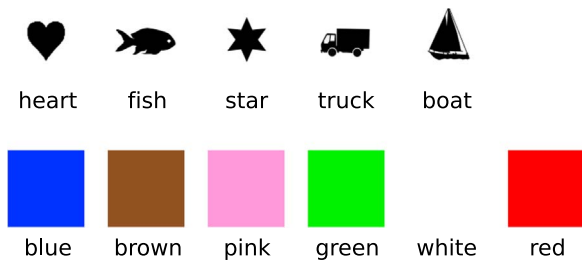
Data for Experiment 2 was collected at New York University, Abu Dhabi. We recruited 30 right handed native speakers of English with no known neurological abnormalities. Data from one female participant were excluded from analysis because she forgot to wear her contact lenses and indicated that she could barely see the stimuli, leaving 16 female and 13 male participants in the analysis (mean age 25.9, range 18–50 years). Because NYU Abu Dhabi is an English speaking university in a country whose main language is not English we paid special care to recruiting only native speakers of English. The majority of the participants, 24 out of the 29, grew up speaking only English, while 4 grew up bilingually and 1 grew up speaking three languages. The protocol was approved by the Institutional Review Board of NYU Abu Dhabi, and subjects provided written consent before beginning the



**Fig. 3.** Experiment 1. *Middle*: Sample trial with two different visual world screens illustrating adjective resolving and noun resolving conditions. *Top*: The analysis of the response to the adjective revealed two temporally distinct clusters in the medial parietal region with an increase in activation when reference was resolved. *Bottom*: The response to the noun also revealed a significant cluster in the medial parietal region with an increase in activation when reference was resolved by the noun. For each cluster, the anatomical image shows the cluster extent (every source that was part of the cluster at some point in time is color-coded with the sum  $F$ - or  $t$ -statistic). A star (\*) indicates that the cluster was significant at  $p \leq .05$ . The time course of activation (for each time point, the average of those sources that were part of the cluster at any time) and total activation (average in the spatio-temporal extent of the cluster) are shown with within-subject standard error (Loftus and Masson, 1994).



**Fig. 4.** Experiment 1, difference between the response to resolving and non-resolving adjectives. Spatio-temporal  $t$ -map thresholded at  $p=.05$  (uncorrected), including all clusters exceeding an extent of 100 sources and 50 ms, displayed in 100 ms time bins. Red/yellow areas indicate more activity in response to reference resolving adjectives, blue areas indicate more activity to non-resolving adjectives.



**Fig. 5.** Experiment 2 stimuli: The colors and shapes used in Experiment 2 with their labels.

experiment.

Because EEG recordings involved wearing an electrode cap including the application of electrolyte gel, we only recorded EEG data for participants that agreed to this more elaborate procedure. This was the case for 14 participants, one of which was excluded because fewer than 50% of the trials remained after artifact rejection. EEG analysis was based on 8 female and 5 male participants (mean age 24.3, range 18–38 years).

### 3.1.2. Design and stimuli

In order to keep the experiment duration under an hour we reduced the number of colors and the number of shapes to 6 and 5, respectively (see Fig. 5). We selected colors and shapes for maximal distinctness in the visual displays, and with the requirement that the shapes completely filled out the area of the aperture that covered them on some trials. For broader validity we also included the object in the middle of the visual world displays as a referent in the adjective resolving condition. Within each cell of the 2 (reference)  $\times$  2 (prediction)  $\times$  3 (location) design we included all possible referents based on permutation of colors and shapes, i.e. 30. Altogether this resulted in 360 trials.

As in Experiment 1, all questions started with *Was the* <color> <shape> and were presented in randomized order to each subject. Displays were systematically different between –reference +prediction trials and +reference –prediction on the one hand, and +reference +prediction and –reference –prediction on the other hand (see Fig. 6): The former contained a hidden object as well as a striped object, whereas the latter contained neither of those. Only in the –reference +prediction condition, where reference had not been resolved at the noun, nouns were followed by a disambiguating *with stripes* or *without stripes*. In addition to these design changes we constructed more natural questions using *next to* as in (5), or asking for absolute location as in (6):

- (5) Was the pink fish next to a boat?
- (6) Was the blue heart with stripes in the middle?

As in Experiment 1, the visual world display and content words were presented for 300 ms with an ISI of 300 ms, whereas short function words like *was*, *the*, *on*, *in* etc. were presented for 200 ms with an ISI of 200 ms. On average, Experiment 2 took 42 min from first to last trial.

### 3.1.3. MEG data acquisition and analysis

MEG data were collected with a 208-channel whole-head axial gradiometer system (Kanazawa Institute of Technology, Nonoichi, Japan) sampling at 1000 Hz with a 200 Hz low-pass filter and a 50 Hz notch filter. Raw data were noise-reduced with the continuously adjusted least squares method (Adachi et al., 2001, see Experiment 1 methods) and low pass filtered at 40 Hz. The data for Experiment 2 were acquired at the NYU Abu Dhabi site. Due to its remote location this facility exhibits lower levels of environmental magnetic noise than the NYU New York site. Consequently, no high pass filter had to be applied (although the electronics of the machine lead to an attenuation

of low frequencies equivalent to a .16 Hz high pass filter). We did not have access to structural MRIs for any of the participants, and used the fsaverage brain for all participants. Apart from those differences, data analysis was analogous to Experiment 1. We analyzed source localized MEG responses to the adjectives with a 2 (resolving vs not resolving reference) by 2 (predicting vs not predicting the noun) by 3 (location of the referent, left, middle or right) ANOVA in the time window spanning both clusters found in Experiment 1 (250–500 ms). Based on the findings of Experiment 1, we restricted our primary search for clusters to the left and right medial parietal regions (i.e., we calculated *p*-values for clusters that were corrected for multiple comparisons across those two regions). In a second, more exploratory stage we included all regions from Experiment 1.

### 3.1.4. EEG data acquisition and analysis

Data were acquired concurrently to the MEG recordings from 31 EEG and 3 EOG electrodes attached to an elastic cap at standard positions in the international 10–20 system (Easycap, Germany) at a sampling rate of 1000 Hz. Impedances were kept below 10 k $\Omega$ .

Raw data were band-pass filtered offline between 0.1 and 40 Hz. We extracted epochs from –100 to 600 ms relative to the onset of the adjectives. Epochs containing artifacts were excluded from further analysis, and individual channels containing noise were interpolated based on the signal in all other electrodes. Epochs were re-referenced to the average of the two mastoid electrodes and baseline corrected using the 100 ms pre-adjective period.

We applied the same cluster permutation method that we also used for MEG source data to the EEG sensor data, in a time window based on the MEG results. Based on the location of the source localized MEG results we restricted the analysis of the EEG data to the posterior half of the electrode cap (see Fig. 8; the following electrodes were included O1, O2, P3, P4, P7, P8, CP1, CP2, CP5, CP6, POZ, PZ, CPZ and the two mastoid electrodes).

EEG data were not used for source localization because individual subject MRIs were not available, making the computation of an exact EEG forward model impossible.

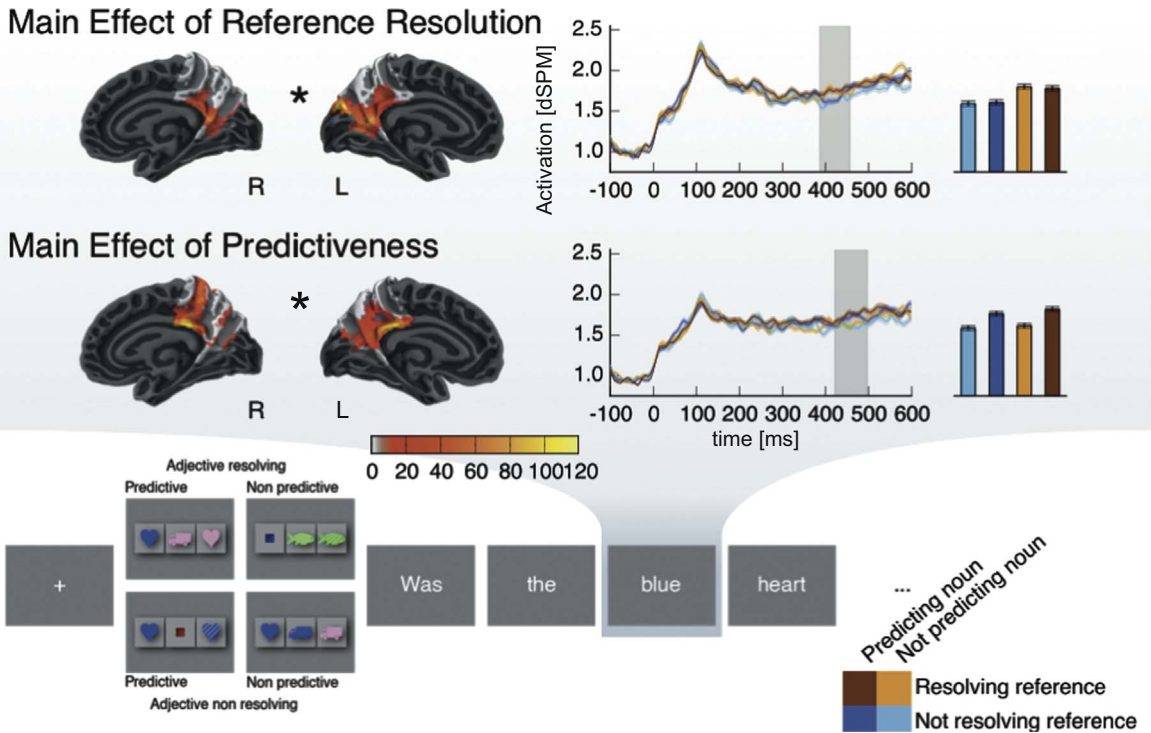
## 3.2. Results and discussion

### 3.2.1. Behavioral results

On average, subjects answered 88.2% of the questions correctly. Participants' individual performance varied between 70.3% and 97.8%. A 2 (reference resolution by adjective or not)  $\times$  2 (adjective predictive of noun or not)  $\times$  3 (location of the referent) ANOVA of the number of incorrect responses indicated a significant interaction of reference and prediction ( $F_{(1,28)} = 7.25$ ,  $p = .012$ ). This interaction was due to more errors in the non-referential predictive condition. This result is likely due to an imbalance in the design, which was optimized to compare brain responses to the color adjectives but not to analyze processing of the subsequent continuations: Referential expressions in the non-referential, predictive conditions were more complex than on the other conditions because they used a third property, the pattern of the referent, for example *the blue heart with stripes*. No effect involving the location of the referent reached significance.

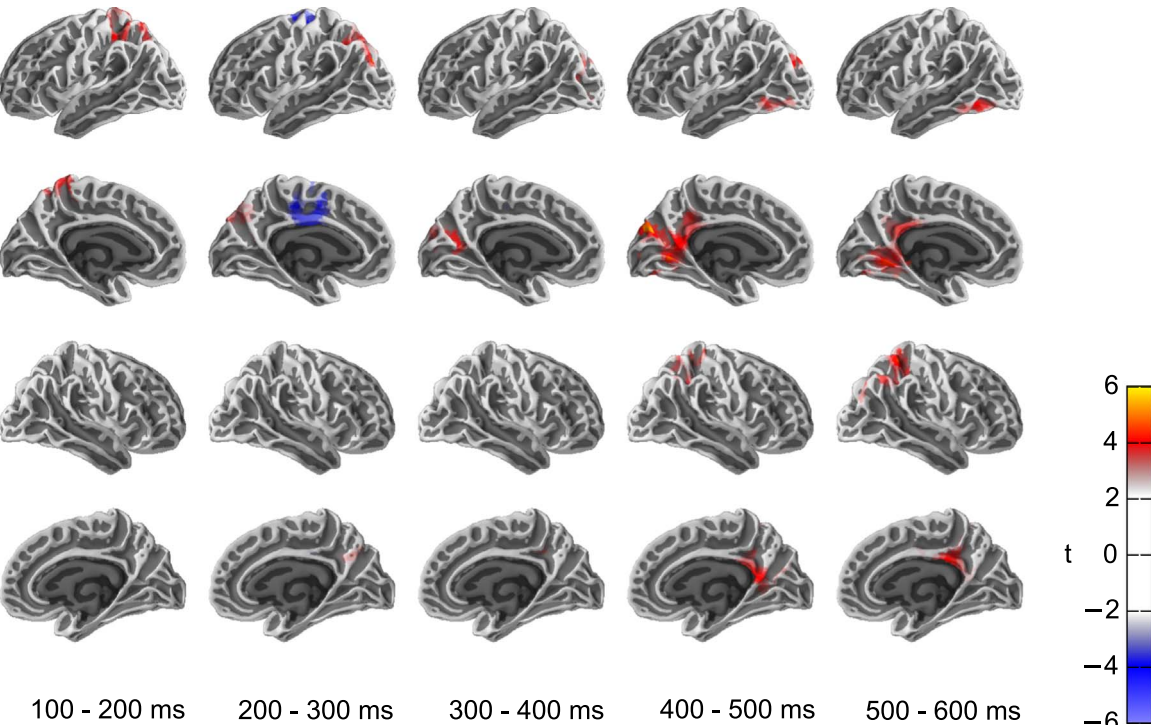
### 3.2.2. MEG results

The analysis of source localized MEG data revealed a cluster with a main effect of reference in the medial parietal region (390–460 ms,  $p = .048$ ), indicating an involvement of the medial parietal lobe in successful reference resolution. The analysis also revealed a cluster with a main effect of prediction (425–495 ms,  $p = .041$ ), suggesting the possibility of a medial parietal involvement in predictive processing in addition to reference resolution. A test including all brain regions used in Experiment 1 did not reveal any additional significant clusters. Fig. 7 illustrates differences between reference resolving and non-resolving adjectives across the whole brain.



**Fig. 6.** Experiment 2. Medial parietal responses to adjectives resolving reference and to adjectives predicting features of the referent. The bottom panel illustrates the four types of visual world displays that determined the possible computations on the highlighted adjective in the subsequent question. In addition to the replication conditions of Experiment 1 (top left and bottom right displays), we used object occlusion to create a condition where reference could be resolved without being able to predict the subsequent noun (top right display) and contrasting patterns to create a condition where reference could not be resolved but the post-adjectival noun could be predicted (bottom left). Details on the cluster plots are analogous to Fig. 3.

There have been concerns that high pass filter settings in the analysis of electrophysiological data can affect the latency of individual responses (see e.g. Tanner et al., 2015). In order to assert that the timing difference in the response to reference resolving adjectives between Experiments 1 and 2 was not due to the difference in filter settings we repeated the analysis of experiment 2 with data high pass filtered at 1 Hz, i.e. the same filter settings that had been used in Experiment 1. Results indicated a cluster with a main effect of reference very similar in distribution and timing to the non high pass filtered analysis (380–435 ms,  $p=.019$ ), but the main effect of predic-



**Fig. 7.** Experiment 2, difference between the response to resolving and non-resolving adjectives (for details see Fig. 4 legend).

tion was not reproduced ( $p \geq .930$ ). These results confirm the reliability of the main effect of reference resolution and suggest that the timing difference between Experiments 1 and 2 was not due to different filter settings.

The MEG results from Experiment 2 confirm a medial parietal involvement in successful reference resolution as well as suggesting a possible involvement in predictive processing. This observation begs the question whether the reference- and prediction related effects are two different properties of the same response, or whether they reflect two different responses. The former could be explained by a response that increases as more information about the referent becomes available. The latter possibility is somewhat favored by the spatial separation between the different clusters, as visual inspection of Fig. 6 suggests that the prediction-related response is somewhat more anterior and the reference-related response somewhat more posterior.

In contrast to Experiment 1, we did not find early effects around 260 ms after adjective onset in Experiment 2. The most salient difference between Experiment 1 and 2 that could explain this difference in timing is the complexity and predictability of the stimuli: in Experiment 1, the pictures were simpler, the questions all followed the same schema, and the reference resolving adjectives could only resolve reference towards the left or the right item in the display. In contrast, Experiment 2 had more complex pictures, questions were more varied and adjectives could resolve reference to any one of the three items. Possibly participants in Experiment 1 noticed this regularity of the stimuli, implicitly or explicitly, and engaged in more strategic processing that allowed them to extract the relevant information, either reference or prediction related, more quickly.

### 3.2.3. EEG results

The analysis of the EEG data in posterior sensors in the 390–460 ms window, based on the MEG results, revealed a significant cluster with a main effect of reference (390–415 ms,  $p=.037$ ). Fig. 8 shows the signal at electrode PZ, the center of the cluster, as well as a topographic difference map in the time window identified by the cluster. No corresponding cluster for the effect of prediction could be found even in the larger 350–500 ms window and before correction for multiple comparisons. The fact that the EEG cluster associated with reference resolution seems to be somewhat shorter than the MEG cluster is most likely due to the specific subgroup of subjects involved: When we repeated the analysis of the MEG source data in this subgroup included in the EEG analysis in the 390–460 ms time window, the medial parietal cluster with main effect of reference also lasted from 390 to 420 ms.

Our result suggests that reference resolution in visual displays can also be tracked with EEG and manifests in a negative going component with maximum at posterior electrodes. Prior EEG studies of referential processing commonly assume that a more positive signal in the N400 time window reflects contextual facilitation (e.g. Burkhardt, 2006). It is

possible that our participants expected the color adjective associated with two items on the display more than the adjective associated with just one item. However, in the context of the experiment both adjectives were equiprobable. Hence our results suggest the possibility that a potential difference in this time window could also reflect referential processing.

## 4. Experiment 3

The referential expressions in Experiments 1 and 2 contained a potentially unnatural element because they always included a color adjective. In a context with a blue heart, a blue tree and a green tree, the incomplete expression *the blue* might bias comprehenders towards the blue tree, because the adjective *blue* is relevant when distinguishing among trees, but not among hearts (Sedivy et al., 1999). Our third experiment avoided this unnaturalness while at the same time replicating our previous finding in a different language, Arabic.

In Arabic, adjectives follow nouns rather than preceding them. Consequently, the reference-resolving *al-jml* (“the-camel”, ALA-LC transliteration) can be contrasted with the visually identical, but non-resolving *al-jml al-azraq* (“the-camel the-blue”), where reference is resolved on the color adjective (see Fig. 10). In this design, referential expressions only contained a color adjective when the noun alone was referentially ambiguous, i.e. questions in the noun-resolving condition were as simple as *Hal al-jml ‘la’ al-ysar* “Was the-camel on the-left?”. This design also avoided the prediction based confound from Experiments 1 and 2: Because nouns in the noun resolving condition were not followed by adjectives, they did not make it possible to predict the subsequent word based on identification of the referent. In addition, questions were balanced such that the noun could never be used to predict the next word.

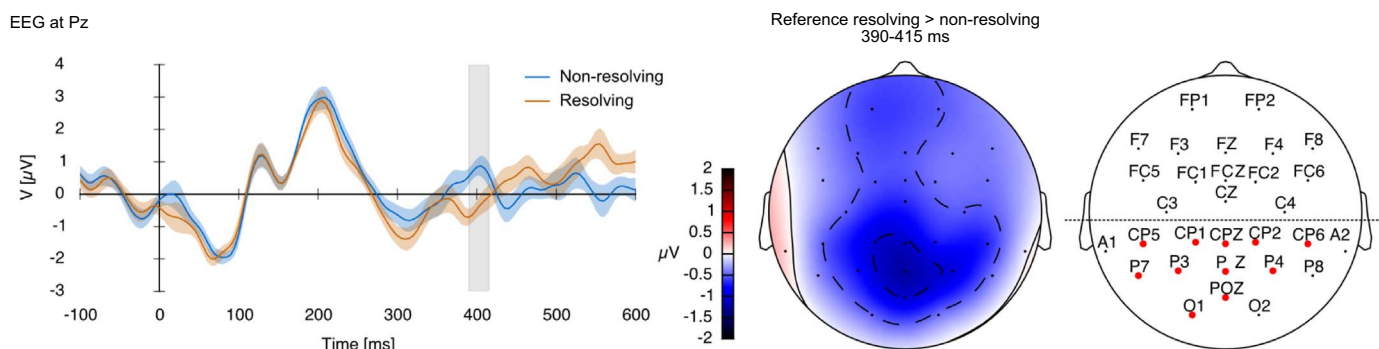
### 4.1. Methods

#### 4.1.1. Participants

Experiment 3 was conducted at the same facilities as Experiment 2 at New York University, Abu Dhabi. We recruited 26 right handed native speakers of Arabic with no known neurological abnormalities. Data from two participants was excluded due to low signal quality, leaving 24 participants (9 female and 15 male, age mean 21.4 years, range 19–24) in the final analysis. Participants came from a variety of linguistic and educational backgrounds, but all had native competence in Modern Standard Arabic. The protocol was approved by the Institutional Review Board of NYU Abu Dhabi, and subjects provided written consent before beginning the experiment.

#### 4.1.2. Design and stimuli

Stimuli for Experiment 3 were based on 8 shapes and 6 colors (see Fig. 9). As in the previous experiments, each possible color/shape



**Fig. 8.** Experiment 2, EEG results. EEG data from Experiment 2 comparing the response to reference resolving and non-resolving adjectives. Left: Signal at sensor PZ with within-subject standard error; gray shading indicates the time window of the significant cluster; Middle: Average difference map in the time window defined by the cluster, 390–415 ms; Right: Electrode map, all electrodes posterior to the dotted line were included in the permutation cluster test. Electrodes that were part of the significant cluster are marked as red dots.

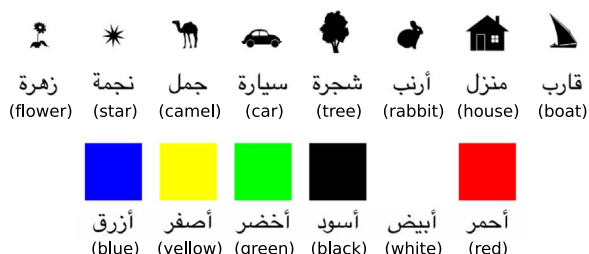


Fig. 9. Experiment 3 Stimuli: Colors and shapes used in Experiment 3.

combination occurred once within each cell of the Reference  $\times$  Location (left/middle/right) design, for a total of 288 trials.

Questions were constructed such that the word following the referential expression was not predictable. The continuation was one of six words (*left, right, on, next-to, between* or *in*), each with equal probability independent of the location of the referent. Questions probed absolute and relative locations, as in examples (7) and (8), respectively:

- (7) هل الزهرة في الوسط؟  
the-middle in the-flower question-particle  
Was the flower in the middle?

- (8) هل النجمة الخضراء بجانب أرنب؟  
rabbit beside the-green the-star question-particle  
Was the green star next to a rabbit?

Presentation rates were analogous to previous experiments. The initial fixation cross was presented for 600 ms, and the visual world display as well as the linguistic stimuli were presented for 300 ms with an ISI of 300 ms. On average, Experiment 3 took 34 min from first to last trial.

MEG data acquisition and analysis were analogous to Experiment 2. A major difference between English and Arabic writing is that the Arabic article is fused with the noun. Consequently, upon perceiving the target word readers in Experiment 3 had to process the direct article, whereas readers in our English studies had already had time to process it since the article had been presented separately 400 ms before the target word. To account for this difference and allow additional time for processing the direct article, we extended the time window for

analysis by 100 ms to 250–600 ms. Based on the findings from the previous experiments the analysis was constrained to the medial parietal region. A 2 (reference) by 3 (location, left/middle/right) ANOVA did not yield significant clusters, but the previous experiments justify applying a one tailed test looking for clusters with an increase in activation for reference resolving over non-resolving nouns.

## 4.2. Results and discussion

### 4.2.1. Behavioral results

Average performance was 87.9% correct answers, with individual participants ranging from 70.8% to 97.9%. The 2 (reference resolution by noun or adjective)  $\times$  3 (referent location) ANOVA indicated a trend in the main effect of location ( $F_{(2,46)} = 3.18, p=.051$ ) due to slightly fewer errors when the referent was on the right ( $M=5.04$ ) than when it was in the middle ( $M=6.29, t_{23} = 2.07, p=.050$ ) and when it was on the left of the display ( $M=6.02, t_{23} = 2.35, p=.028$ ). This suggests the possibility that participants exhibited an attentional bias towards the right side of the visual display, leading to improved performance. While an attentional bias towards the right would be consistent with Arabic spelling going from right to left, we note that we did not find a corresponding bias towards the left side in native English speakers.

### 4.2.2. MEG results

The analysis of the response to nouns with a related measures *t*-test for an increase in activation associated with reference resolution yielded a significant cluster (455–600 ms,  $p=.026$ , see Fig. 10). Pairwise tests of the total cluster activation confirmed that each referential sub-condition was associated with more activation than the non-referential condition. Inclusion of the additional search regions included in Experiment 1 revealed no additional significant clusters. Fig. 11 illustrates differences between reference resolving and non-resolving nouns across the whole brain.

Experiment 3 thus replicated our previous results in Arabic. This is in line with our interpretation that the effect is related to an aspect of language comprehension that is shared across languages, and does not depend on the specific writing system (i.e., English or Arabic). In addition, this result confirms an involvement of the medial parietal lobe in reference resolution when the critical expression is a noun not preceded by an adjective, and not predicting the subsequent word. Finally, the sample of participants that provided the data for Experiment 3 consists of a more homogeneous population in terms of age than Experiments 1 and 2. The replication thus also indicates

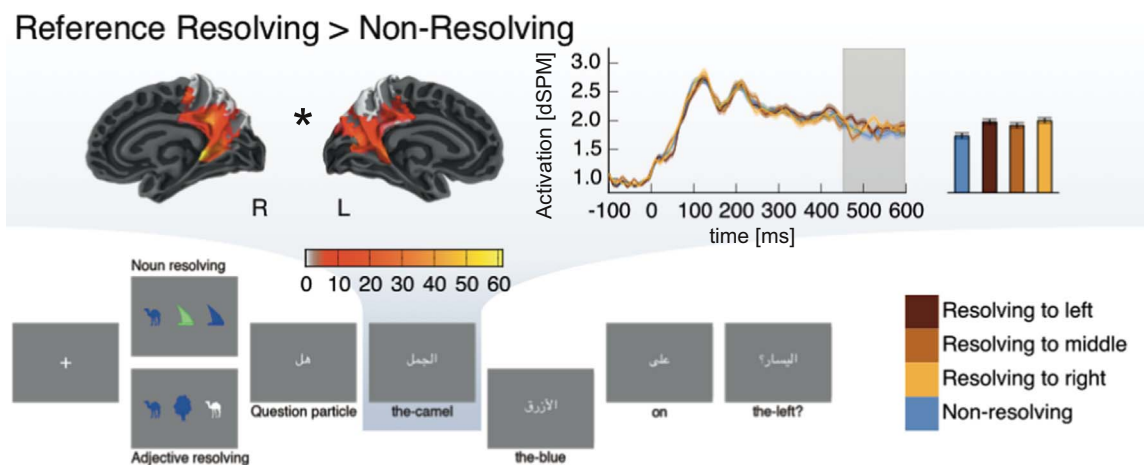
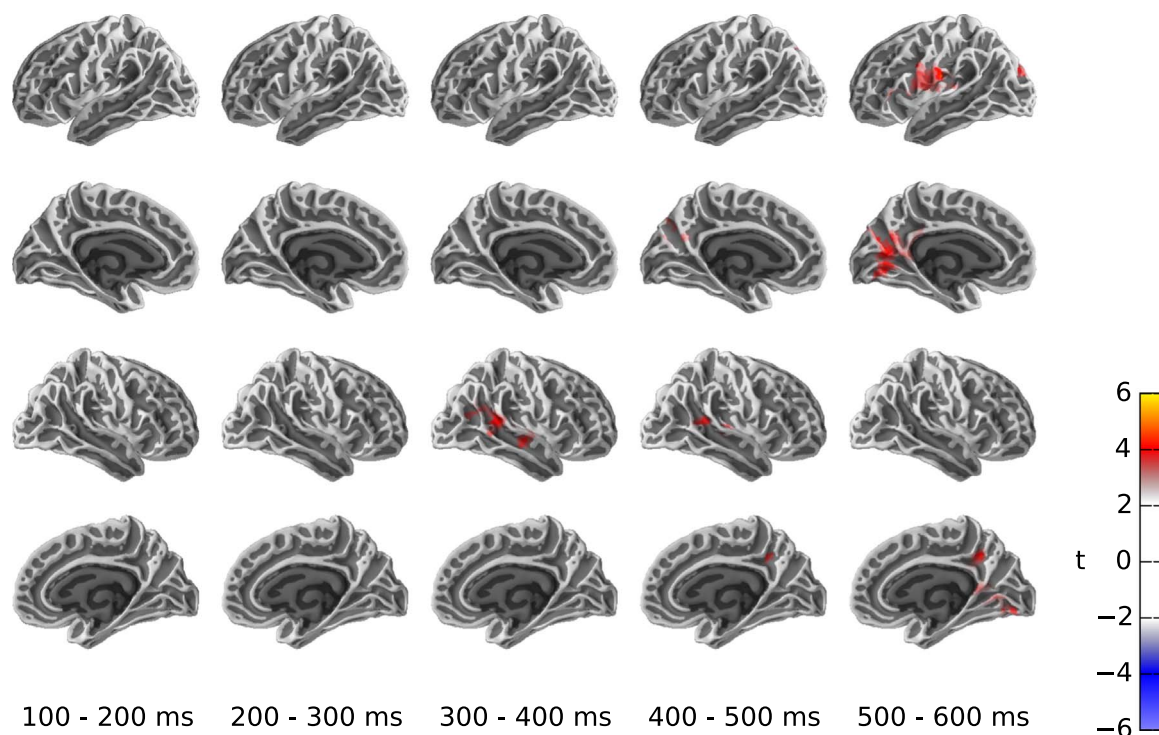


Fig. 10. Experiment 3: Medial parietal responses to reference resolving nouns depending on the location of the referent in Arabic. To dissociate prediction from reference resolution, Experiment 3 took advantage of the fact that in Arabic, adjectival modifiers follow the noun. The highlighted noun described either a unique item in the just viewed display, allowing reference resolution, or two distinctly colored items, not allowing reference resolution. In the latter case, a subsequent color adjective resolved the reference (the three levels on the factor location of the referent are not shown). The adjective screen (“the-blue”) is shown on the level of the adjective resolving display because it was only presented in that condition. The analysis in the medial parietal region revealed a cluster of increased activation following reference resolving nouns. Details on the cluster plots are analogous to Fig. 3.



**Fig. 11.** Experiment 3, difference between the response to resolving and non-resolving nouns (for details see Fig. 4 legend).

that the effect is not dependent on a specific age composition.

## 5. General discussion

Across three experiments, we observed an increase in brain activity localized to the medial parietal lobe when a word provided enough information for reference resolution. The effect was replicated across different subject populations and different experimental paradigms. While Experiment 1 confounded referential and predictive properties of the target word, the crossed design of Experiment 2 confirmed a medial parietal response associated with reference resolution, at the same time suggesting the possibility of a response associated with prediction. Experiment 3 replicated the response to reference resolution in Arabic, in a pragmatically unmarked context and without a prediction confound. Interestingly, we found medial parietal activation in response to reference resolving words regardless of whether they constituted a complete noun phrase (*the house*) or whether they were part of an ongoing, incomplete noun phrase (*the green ...*). This observation supports the notion that the medial parietal lobe is engaged when reference can be resolved at incomplete as well as complete phrases, independently of the specific linguistic construction.

Experiment 1 suggested an early onset for the response to reference resolving adjectives (255 ms), but this was not replicated in the analysis of the nouns in Experiment 1, which suggested an onset around 340 ms, nor in the follow up experiments which suggested an onset around 390 ms for English adjectives and 455 ms for Arabic nouns. The delay for Arabic might be due to the fact that in Arabic, articles are fused with the words to which they apply, i.e. readers in English had a head start by being presented with the definite article *the* separately before the adjective, whereas readers of Arabic had to process the definite article upon presentation of the target word. A possible explanation for the early response in Experiment 1 is the predictability of the stimulus material. In Experiment 1, all questions were constructed according to the same schema by filling in a color, two shapes and a relation (*opposite* or *beside*; see examples 2 and 3) and referents of reference resolving adjectives could only be in two possible locations, on the left and the right side of the display. This might have

lead readers to engage in strategic processing which allowed them to extract referential or predictive information more quickly than in Experiments 2 and 3, where stimuli were more varied. Overall, the timing of our results is consistent with EEG studies of referential ambiguity which place the time point at which referential ambiguity is registered in the 300–400 ms range (e.g. Van Berkum et al., 1999; Nieuwland and Van Berkum, 2006).

Taken together, our results suggest that this brain response is a promising candidate for a neural measure of reference resolution. As with any brain response, 3 experiments are not sufficient to definitely establish the functional significance of the response. However, we have narrowed the possibilities considerably. First, our results have shown that the response depends on reference resolution regardless of whether it allows predicting the referent's shape and upcoming linguistic material. Second, our results suggest that the response does not depend on a complete linguistic phrase, as it occurred at adjectives, i.e. in unfinished noun phrases, as well as nouns which concluded a noun phrase. And third, we have obtained a comparable response pattern in two different languages, English and Arabic.

An open question is what level of representation the medial parietal response is related to, a question that is tied to the question on the role of the medial parietal lobe in cognition in general. Based on our results there is a continuum of possibilities ranging from a visual short term buffer to more abstract, discourse level representations. On the one hand, our participants needed a representation of the referential domain, the visual display, which we assume to reside in visual short term memory. Reference resolution amounted to identifying and selecting an entity in that representation. An increase in activation could thus reflect reactivation of the representation of the referent in a visual short term store. On the other hand, participants had to construct a model of the proposition that is asked in the question as an abstract, linguistic representation (e.g. of the kind described by Kintsch (1988)). In this representation, increased activation could reflect the addition of a new entity to the proposition or its integration with prior context. EEG data acquired concurrently with Experiment 3 (Brodbeck et al., 2015) revealed an effect of reference resolution that reflects the location of the referent, bearing a hallmark of spatial

memory. This is compatible with a model in which the representation of the referent is connected to a spatial memory system. With an onset around 333 ms, this effect preceded the medial parietal activity found in the source localized MEG data, which started around 390 ms. The medial parietal result described here is characterized by a main effect of reference resolution, unaffected by the location of the referent, and thus reflects a generalization over the EEG effect. As such it is compatible with a mechanism involving spatial memory as well as a modality-general, discourse level process. The extant literature also seems to provide some support for both possibilities:

On the one hand, precuneus (Cavanna and Trimble, 2006) as well as retrosplenial cortex (Vann et al., 2009), both regions of the medial parietal lobe, have been implicated in processing visuospatial representations. fMRI studies have found medial parietal lobe activation associated with making spatial judgements about visual scenes (Committeri et al., 2004; Galati et al., 2010) as well as when answering questions about spatial relations in visually presented (Wallentin et al., 2006) or described (Wallentin et al., 2008) scenes. The precuneus is also one of many brain regions that contain so called eye fields, which exhibit activity related to eye movements and together are thought to form a distributed network for eye movement control (Lynch and Tian, 2006) and spatial attention (e.g. Corbetta and Shulman, 2011). It is thus plausible that activity in this region reflects coding of the spatial location of the referent or an attentional shift to its position, and might even be involved in planning of a saccade towards the referent. Notably, however, a meta-analysis of visual imagery did not indicate significant involvement of the medial parietal lobe (McNorgan, 2012), suggesting that visual imagery by itself is unlikely to be the driving factor for our result.

On the other hand, the medial parietal lobe is implicated in a number of paradigms related to discourse comprehension that do not have an explicit visual component, including referentially ambiguous pronouns (Nieuwland et al., 2007b), coherent language (Ferstl et al., 2008) and tracking conjoined discourse referents (Boiteau et al., 2014). More generally, the medial parietal lobe has been identified as supporting representations of context in “binding of item and context” models of memory. Ranganath and Ritchey (2012) argued that the ability to place an item in the context of other representations is tied to posterior medial temporal/retrosplenial regions, which are functionally connected to much of the medial parietal lobe (Kahn et al., 2008). They even suggested that the posterior/medial parietal system might be involved in representing situation models in language processing as described by Kintsch (1988). This claim is highly compatible with an involvement in reference resolution, which involves locating an item that is embedded in a situation model and including it in the proposition that is under construction.

Those two possibilities are also not mutually incompatible. It is possible that visual manipulations engage medial parietal cortex precisely because they involve visual and spatial situation models. Or, conversely, that the capacity to represent abstract situation models developed out of the capacity to represent spatial models, as has been argued for the hippocampus in regard to spatial and episodic memory (e.g. Allen and Fortin, 2013). A similar parallel has been suggested to account for the observation that superior parietal areas are involved in processing discourses with multiple referents (Almor et al., 2007; Boiteau et al., 2014). This dual function hypothesis is also consistent with lesion evidence, since lesions of retrosplenial cortex can lead to severe general episodic memory deficits, as well as topographical disorientation, a navigation deficit despite intact landmark recognition (Maguire, 2001). Nevertheless, an important goal for future research should be to test whether the activity we observed is tied to a modality-specific, visuospatial representations or not.

Together with results cited in the Introduction, our results suggest a cortical division of labor during reference resolution. While our results associate parietal cortex with tracking referents on a basic level, in line with the results of Boiteau et al. (2014) and Almor et al. (2007), frontal

lobe activity seems to be associated with higher order decision making processes in non-trivial referential situations such as referential ambiguity (Nieuwland et al., 2007b; McMillan et al., 2012).

An important distinction that might warrant future attention is the mechanism of reference resolution and its relationship to processing of the sets of referential candidates. One way to model reference resolution is through sets: Referential processing starts with a set of all entities in the referential domain, and any information is used to narrow down this set. For example, if the current referential candidates is a set of yellow and blue objects, the adjective *blue* constrains the set of referential candidates to the subset of blue objects. The cognitive act of reference resolution can be seen as pruning a set of potential referents until a unique object is left. In the context of our paradigm, the set of potential referents is initially the set of the three objects in the display, which is then pruned with content words like color adjectives and shape-describing nouns until a single item remains in the set, the referent. In terms of this model, the main contrasts in Experiments 1–3 compared reduction of the candidates from 3 to 1 with a reduction from 3 to 2 (e.g., *blue* in the context with one vs two blue objects). The analysis of the noun in Experiment 1 compared reduction from 2 to 1 with no reduction at all (when the preceding adjective had already narrowed the set to 1). Consequently, the medial parietal effect we obtained could also be related to a brain response that scales with the number of referents that could be excluded (excluding 2 vs 1, or 1 vs 0). Future work should address this important issue.

A further issue concerns the properties of the referential domain. Visual world studies are sometimes criticized for involving artificially small referential domains. This criticism could be leveled against our results in particular. While some real world referential domains are certainly much larger, there are also situations where speakers and listeners refer to items from a constrained set. A good example that comes close to the kinds of domains we studied might be deliberations about articles of clothing, as when someone needs help deciding between *the green sweater, the green shirt or the white shirt*. Such situations might or might not recruit the same cognitive mechanism as more out-of-the-blue referential situations with larger domains. In particular, the decision to accept a specific item as the referent might be simplified when the given property (e.g. a color adjective) only has to be compared with three items from short term memory as opposed to a larger domain from long term memory. As most research does, we thus identified a specific example of the phenomenon we intend to study, one that lends itself to a particular operationalization, and future research will have to show whether our results generalize to different scenarios, or whether different cognitive processes are involved in referential processing depending on the nature of the referential situation.

In addition to effects of referential processing, Experiment 2 suggested the possibility that the medial parietal lobe also responded to the predictiveness of the color adjective. We observed a trend towards an increase in activation when the adjective made it possible to infer the shape of the referent and thus predict the noun. A possible commonality between this kind of prediction and reference resolution is that new information about the referent becomes available; in the case of reference resolution this is identity information, in the case of prediction it is information about the shape of the referent. The activity could reflect integration of new information with the model of the unfolding proposition. While it might seem counter-intuitive to claim that we integrate information about an entity without knowing which entity it is, i.e., before reference resolution, this is in fact a common phenomenon with so-called attributive interpretations of definite descriptions, as when *the murderer of Smith is crazy* is interpreted to mean “whoever killed Smith (I do not know who it was) is crazy”.

It is in order to point out some methodological limitations. While visual world studies typically employ spoken language stimuli, emphasizing naturalness of the stimulus material, we opted for visual stimuli. This is in line with a long tradition in MEG/EEG research taking

advantage of the more highly controllable nature of visual language stimuli. Important for MEG/EEG, the temporal characteristics of written stimuli, presented word by word, can be strictly controlled, whereas natural speech stimuli introduce more temporal variability into the signal. As a consequence of this, written stimuli are studied more frequently with electrophysiological methods, and more is known about the time course of written language processing. Using written stimuli thus allowed us to compare our results to a relatively well known time course of written word perception (e.g. Grainger and Holcomb, 2009). As with other paradigms, we expect that future work will explicitly compare responses obtained under visual and auditory stimulation to address questions about the relationship between processing written and spoken language. There is at least some evidence indicating that referential processing is similar for written and spoken language from the observation of NRef responses to written (Van Berkum et al., 1999) as well as spoken material (Nieuwland et al., 2007a).

Finally, while visual world studies described incremental reference resolution in speech stimuli, we believe that our results (along with those presented in Brodbeck et al. (2015)) are the first to show incremental referential processing in response to adjectives in written language.

## 6. Conclusions

In sum, this work identified a neural correlate of successful reference resolution localized to the medial parietal lobe. Across three experiments, we demonstrated that medial parietal cortex responds more to reference resolving than non-resolving stimuli. We were able to characterize the temporal profile of this activity, placing it in a mid-latency time-window, at 350–500 ms after the onset of the resolving word, with the details depending on the specifics of the stimuli, and the possibility that highly constrained stimuli can be associated with an even earlier response.

Taking into account previous results we can start to map out a sequence of events involved in reference resolution. In previous work we identified an EEG response sensitive to the location of the referent, starting around 333 ms (Brodbeck et al., 2015). Since these data were collected together with the MEG data from Experiment 2 described here, we can directly compare the time course of the two effects. This comparison suggests that the location-specific EEG response preceded the medial parietal response to reference resolution, which started at 390 ms. Together, these results suggest that an attentional shift towards the location of the referent slightly precedes a medial parietal response which reflects reference resolution independently of the location of the referent.

Together our results thus constitute a promising starting point for multiple lines of research, including not only investigations of referential language processing and its brain basis, but also work aimed at understanding the place of reference resolution within more general models of the neural basis of language, which have yet to incorporate this key aspect of language processing (e.g. Friederici, 2011; Hagoort and Indefrey, 2014).

## Acknowledgements

This research was supported by the National Science Foundation Grant BCS-1221723 (LP) and grant G1001 from the NYUAD Institute, New York University Abu Dhabi (LP). We thank Laura Gwilliams for help with the data collection for Experiment 2.

## References

Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., Ochiai, M., 2001. Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. *IEEE Transactions on Applied Superconductivity* 11 (1), 669–672.

- Allen, T.A., Fortin, N.J., 2013. The evolution of episodic memory. *Proceedings of the National Academy of Sciences* 110, 10379–10386.
- Allopenna, P.D., Magnuson, J.S., Tanenhaus, M.K., 1998. Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language* 38 (4), 419–439.
- Almor, A., Smith, D.V., Bonilha, L., Fridriksson, J., Rorden, C., 2007. What is in a name? Spatial brain circuits are used to track discourse references. *Neuroreport* 18 (12), 1215–1219.
- Altmann, G.T.M., Kamide, Y., 1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73 (3), 247–264.
- Altmann, G.T.M., Kamide, Y., 2007. The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language* 57 (4), 502–518.
- Anderson, J.E., Holcomb, P.J., 2005. An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain and Language* 94 (2), 200–216.
- Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex* 19 (12), 2767–2796.
- Boiteau, T.W., Bowers, E., Nair, V.A., Almor, A., 2014. The neural representation of plural discourse entities. *Brain and Language* 137, 130–141.
- Boudewyn, M.A., Long, D.L., Traxler, M.J., Lesh, T.A., Dave, S., Mangun, G.R., Carter, C.S., Swaab, T.Y., 2015. Sensitivity to Referential Ambiguity in Discourse: The Role of Attention, Working Memory, and Verbal Ability. *Journal of Cognitive Neuroscience* 27 (12), 2309–2323.
- Brodbeck, C., Gwilliams, L., Pykkänen, L., 2015. EEG can track the time course of successful reference resolution in small visual worlds. *Frontiers in psychology* 6, 1787.
- Burkhardt, P., 2006. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain Lang* 98, 159–168.
- Burkhardt, P., 2007. The P600 reflects cost of new information in discourse memory. *NeuroReport* 18 (17), 1851–1854.
- Cavanna, A.E., Trimble, M.R., 2006. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129 (3), 564–583.
- Chambers, C.G., Tanenhaus, M.K., Eberhard, K.M., Filip, H., Carlson, G.N., 2002. Circumscribing Referential Domains during Real-Time Language Comprehension. *Journal of Memory and Language* 47 (1), 30–49.
- Committeri, G., Galati, G., Paradis, A.-L., Pizzamiglio, L., Berthoz, A., LeBihan, D., 2004. Reference Frames for Spatial Cognition: Different Brain Areas are Involved in Viewer-, Object-, and Landmark-Centered Judgments About Object Location. *Journal of Cognitive Neuroscience* 16 (9), 1517–1535.
- Corbetta, M., Shulman, G.L., 2011. Spatial Neglect and Attention Networks. *Annual Review of Neuroscience* 34 (1), 569–599.
- Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26 (1), 55–67.
- Eberhard, K.M., Spivey-Knowlton, M.J., Sedivy, J.C., Tanenhaus, M.K., 1995. Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research* 24 (6), 409–436.
- Ferstl, E.C., Neumann, J., Bogler, C., von Cramon, D.Y., 2008. The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping* 29 (5), 581–593.
- Fischl, B., Sereno, M.I., Tootell, R.B.H., Dale, A.M., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping* 8 (4), 272–284.
- Friederici, A.D., 2011. The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews* 91 (4), 1357–1392.
- Galati, G., Pelle, G., Berthoz, A., Committeri, G., 2010. Multiple reference frames used by the human brain for spatial perception and memory. *Experimental Brain Research* 206 (2), 109–120.
- Grainger, J., Holcomb, P.J., 2009. Watching the Word Go by: On the Time-course of Component Processes in Visual Word Recognition. *Language and Linguistics Compass* 3 (1), 128–156.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M.S., 2013. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* 7, 267.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG data. *NeuroImage* 86, 446–460.
- Hagoort, P., Indefrey, P., 2014. The Neurobiology of Language Beyond Single Words. *Annual Review of Neuroscience* 37 (1), 347–362.
- Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & Biological Engineering & Computing* 32 (1), 35–42.
- Hauk, O., Wakeman, D.G., Henson, R.N., 2011. Comparison of noise-normalized minimum norm estimates for MEG analysis using multiple resolution metrics. *NeuroImage* 54 (3), 1966–1974.
- Huetttig, F., Rommers, J., Meyer, A.S., 2011. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica* 137 (2), 151–171.
- Hutton, S.B., 2008. Cognitive control of saccadic eye movements. *Brain and Cognition* 68 (3), 327–340.
- Kahn, I., Andrews-Hanna, J.R., Vincent, J.L., Snyder, A.Z., Buckner, R.L., 2008. Distinct Cortical Anatomy Linked to Subregions of the Medial Temporal Lobe Revealed by Intrinsic Functional Connectivity. *Journal of Neurophysiology* 100 (1), 129–139.

- Kamide, Y., Altmann, G.T.M., Haywood, S.L., 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language* 49 (1), 133–156.
- Kintsch, W., 1988. The Role of Knowledge in Discourse Comprehension - a Construction Integration Model. *Psychol Rev* 95, 163–182.
- Lambrecht, K. (1994). Information structure and sentence form: Topics, focus, and the mental representations of discourse referents, volume xvi of Cambridge studies in linguistics. Cambridge University Press.
- Ledoux, K., Gordon, P.C., Camblin, C.C., Swaab, T.Y., 2007. Coreference and lexical repetition: Mechanisms of discourse integration. *Memory & Cognition* 35 (4), 801–815.
- Loftus, G.R., Masson, M.E.J., 1994. Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review* 1 (4), 476–490.
- Luck, S.J., Vogel, E.K., 1997. The capacity of visual working memory for features and conjunctions. *Nature* 390 (6657), 279–281.
- Lynch, J., Tian, J.-R. 2006. Cortico-cortical networks and cortico-subcortical loops for the higher control of eye movements. In: *Progress in Brain Research*, volume 151, Elsevier, pages 461–501.
- Maguire, E.A., 2001. The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. *Scandinavian journal of psychology* 42 (3), 225–238.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* 164 (1), 177–190.
- McMillan, C.T., Clark, R., Gunawardena, D., Ryant, N., Grossman, M., 2012. fMRI evidence for strategic decision-making during resolution of pronoun reference. *Neuropsychologia* 50 (5), 674–687.
- McNorgan, C., 2012. A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. *Frontiers in Human Neuroscience*, 6.
- Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping* 15 (1), 1–25.
- Nieuwland, M.S., Otten, M., Van Berkum, J.J.A., 2007a. Who are You Talking About? Tracking Discourse-level Referential Processing with Event-related Brain Potentials. *Journal of Cognitive Neuroscience* 19 (2), 228–236.
- Nieuwland, M.S., Petersson, K.M., Van Berkum, J.J.A., 2007b. On sense and reference: examining the functional neuroanatomy of referential processing. *NeuroImage* 37 (3), 993–1004.
- Nieuwland, M.S., Van Berkum, J.J.A., 2006. Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research* 1118 (1), 155–167.
- Nieuwland, M.S., Van Berkum, J.J.A., 2008. The Neurocognition of Referential Ambiguity in Language Comprehension. *Language and Linguistics Compass* 2 (4), 603–630.
- Ranganath, C., Ritchey, M., 2012. Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience* 13 (10), 713–726.
- Sedivy, J.C., Tanenhaus, K., Chambers, M., C.G., Carlson, G.N., 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71(2):109–147.
- Spivey, M.J., Tanenhaus, M.K., Eberhard, K.M., Sedivy, J.C., 2002. Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology* 45 (4), 447–481.
- Tanenhaus, M.K., Brown-Schmidt, S., 2008. Language processing in the natural world. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1493), 1105–1122.
- Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., Sedivy, J.C., 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268 (5217), 1632–1634.
- Tanner, D., Morgan-Short, K., Luck, S.J., 2015. How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition: High-pass filtering and artifactual ERP effects. *Psychophysiology* 52 (8), 997–1009.
- Van Berkum, J.J.A., Brown, C.M., Hagoort, P., 1999. Early Referential Context Effects in Sentence Processing: Evidence from Event-Related Brain Potentials. *Journal of Memory and Language* 41 (2), 147–182.
- Van Berkum, J.J.A., Brown, C.M., Hagoort, P., Zwitterlood, P., 2003. Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology* 40 (2), 235–248.
- van Deemter, K., 2016. Computational models of referring: a study in cognitive science. The MIT Press.
- Van Essen, D.C., 2005. A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *NeuroImage* 28 (3), 635–662.
- Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., McIsaac, H., 1991. Fractionating the Word Repetition Effect with Event-Related Potentials. *Journal of Cognitive Neuroscience* 3 (2), 131–150.
- Vann, S.D., Aggleton, J.P., Maguire, E.A., 2009. What does the retrosplenial cortex do? *Nature Reviews Neuroscience* 10 (11), 792–802.
- Wallentin, M., Roepstorff, A., Glover, R., Burgess, N., 2006. Parallel memory systems for talking about location and age in precuneus, caudate and Broca's region. *NeuroImage* 32 (4), 1850–1864.
- Wallentin, M., Weed, E., Østergaard, L., Mouridsen, K., Roepstorff, A., 2008. Accessing the mental space-Spatial working memory processes for language and vision overlap in precuneus. *Human Brain Mapping* 29 (5), 524–532.
- Zelinsky, G.J., Murphy, G.L., 2000. Synchronizing Visual and Language Processing: An Effect of Object Name Length on Eye Movements. *Psychological Science* 11 (2), 125–131.