1 Language in context: Characterizing the comprehension of 2 referential expressions with MEG 3

4 Christian Brodbeck^{1,2} and Liina Pykkänen^{1,2,3}

5 ¹Department of Psychology, New York University, New York, NY, 10003, USA

6 ²NYU Abu Dhabi Institute, Abu Dhabi, United Arab Emirates

7 ³Department of Linguistics, New York University, New York, NY, 10003, USA
8

9 Contact:

10 Christian Brodbeck

11 New York University, Department of Psychology

12 6 Washington Place

13 New York, NY 10003

14 U.S.A.

15 Telephone: +1 917 353 1422

16 christianbrodbeck@nyu.edu

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 medial parietal lobe. In each trial, participants saw a simple visual display containing three object 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 expression 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 suggest avenues for future research.

Keywords: MEG, EEG, sentence comprehension, reference resolution, visual short-term memory

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 & 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, Spivey-Knowlton, Eberhard, & Sedivy, 1995), takes into account a wide array of extralinguistic sources of information (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Altmann & Kamide, 2007) and can even affect syntactic parsing decisions (Tanenhaus et al., 1995; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). Most of this evidence has come from studies using the so-called visual world paradigm, in which participants’s 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 Brown-Schmidt 2008; Huettig, Rommers, and Meyer 2011). Studies that focused directly on reference 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, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Sedivy, K. Tanenhaus, Chambers, & Carlson, 1999). For example, when participants were asked to *Touch the starred yellow square* in a context with only one starred item, they moved their

67 eyes to the relevant item on average 250 ms after the end of the word *starred*, significantly earlier
68 than in contexts where more than one object had stars. However, eye movements constitute an
69 indirect measure and do not necessarily reflect reference resolution to a unique item. When the
70 first syllable of an utterance is compatible with two different continuations (e.g., *the beetle* vs *the*
71 *beaker*), addressees distribute their fixations over both items (Allopenna, Magnuson, & Tanenhaus,
72 1998). This suggests that eye movements reflect attentional processes, guided by even partial word
73 information, and before reference can be positively resolved.

74 Neural measures can provide complementary information to improve our understanding of
75 the computational stages of reference resolution. In particular, electrophysiological measures can
76 provide temporally precise signals, making it possible to measure the response to individual words
77 in coherent language stimuli. A repeated finding from event related potential (ERP) studies
78 is an N400 reduction to expressions resolving reference to an entity introduced in the previous
79 sentence in a non-marked way (Burkhardt, 2006; Ledoux, Gordon, Camblin, & Swaab, 2007).
80 Referential context can influence this response, suggesting that referential processing is occurring
81 in the relevant time window between 250 and 500 ms (Ledoux et al., 2007). A later positive
82 component has been associated with retrieval and updating when a prior referent is mentioned
83 again (Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991). However, this observation is
84 complicated by the fact that a similar response has also been associated with the introduction of a
85 new discourse referent, when compared with reference to an existing one (Burkhardt, 2006, 2007).
86 Yet another study that manipulated referential status through the article (*Kathy sat nervously in*
87 *the cab on her way to the airport. A/The cab...*) found no late component in either direction
88 (Anderson & Holcomb, 2005). Together these results suggest that the late ERP component is
89 sensitive to multiple factors and not yet interpretable as a direct measure of referential status per se.
90 Another brain measures, which has achieve a higher degree of functional specificity is a component
91 related to referential ambiguity. A group of EEG studies have found that referentially ambiguous
92 expressions evoke a sustained frontal negative-going event-related potential when compared to
93 unambiguous controls (reviewed by Nieuwland and Van Berkum, 2008). This includes determiner-

94 noun phrases like *the girl* after a story context that introduced two girls vs one girl (Van Berkum,
95 Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitterlood, 2003; Nieuwland, Otten,
96 & Van Berkum, 2007; Boudewyn et al., 2015) as well as pronouns matching two vs one previously
97 introduced characters (e.g., *he* in *Ronald told Frank that he ...* can refer to *Ronald* as well as
98 *Frank*, whereas *he* in *Ronald told Emily that he ...* can only refer to *Ronald*) (Nieuwland &
99 Van Berkum, 2006).

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

113 Our investigation had the goal of testing for activity in these regions that is immediately as-
114 sociated with successful reference resolution when higher order decision making requirements are
115 minimal. A challenge for studying successful reference resolution is finding a suitable control condi-
116 tion that does not involve creation of new discourse referents while at the same time not introducing
117 another contrast like referential failure or referential ambiguity. To this end we adapted the visual
118 world paradigm that has been used for studying incremental reference resolution (Eberhard et al.,
119 1995, see above) for MEG/EEG in three experiments, each with the same basic trial structure (see
120 Figure 3 for Experiment 1). In each trial, participants first saw a simple visual world display and

121 then read a question about that display, presented word by word. Participants were required to
122 answer each question with a yes/no button press, encouraging them to process the questions natu-
123 rally with the goal of comprehending them and without drawing undue attention to the referential
124 aspect of the task. We analyzed the neural response to simple referential expressions like *the blue*
125 *heart*, comparing the same expressions in different referential contexts. In the example in Figure
126 3 the adjective *blue* could resolve reference in a context with one blue item, but not in a context
127 with two blue items. This contrast isolates reference resolution with minimal overt ambiguity.
128 Based on EEG data collected concurrently with the MEG data reported here for Experiment 2,
129 we previously reported that evoked potentials to reference resolving words reflect the location of
130 the referent on the display around 333 ms after adjective onset, confirming that the manipulation
131 is effective in engaging referential processing at the adjective (Brodbeck, Gwilliams, & Pylkkänen,
132 2015). MEG allowed us to measure brain activity associated with the processing of individual
133 words with high temporal precision. This allowed us to track neural activity occurring during the
134 processing of specific words in referential expressions, in contrast to fMRI which does not allow
135 attributing activity temporally to individual words in a sentence.

136 While the fMRI study on referential ambiguity (Nieuwland, Petersson, & Van Berkum, 2007,
137 see above) provides us with a broad set of regions of interest, there are also other relevant prior
138 results. First, another fMRI study found that discourses involving two conjoined referents com-
139 pared to discourses with one or two singular referents engaged medial and superior/lateral parietal
140 regions (Boiteau, Bowers, Nair, & Almor, 2014). This result suggests that parietal cortex is in-
141 volved in the representation of discourse referents because it is sensitive to the manner in which
142 multiple discourse entities are introduced. Second, reference resolution is a fundamental compo-
143 nent of processing coherent language, because coherence critically depends on repeated reference
144 to the same entities. A meta-analysis found that the processing of coherent language is associated
145 with increased activity in medial parietal, medial frontal and bilateral temporal areas (Ferstl, Neu-
146 mann, Bogler, & von Cramon, 2008). Together, these results highlight the parietal lobe as possibly
147 relevant for referential language processing, since it is the only region that was reliably affected by

all three contrasts. Almor, Smith, Bonilha, Fridriksson, and Rorden (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.

169 2.1 Methods

170 2.1.1 Participants

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

176 2.1.2 Design and stimuli

177 To construct the visual world displays, we identified 7 common, monosyllabic color adjectives, and
178 7 nouns matched to the adjectives for frequency and the number of characters (see Figure 1). All
179 displays contained three objects in a horizontal array and were constructed using two colors and
180 two shapes (see Figure 3). In each display, one of the three objects had a unique color and another
181 object had a unique shape. The object with the unique color was always located on the outside of
182 the display (either left or right).

Figure 1: Experiment 1 stimuli: The 7 color and the 7 shapes used in Experiment 1, labeled with the words used in the questions.

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

185 (1) Was the <color> <shape> beside a <shape>?

186 (2) Was the <color> <shape> opposite a <shape>?

187 (3) Was the green key beside a bomb?

188 (4) Was the green key opposite a bomb?

189 The relations, *beside* and *opposite*, were explained to participants before the experiment (*opposite*
190 referring to the opposition of the two outside items). In half the questions, the color adjective
191 identified a single item, whereas in the other half the display contained two items of the color and
192 the shape described by the noun allowed identification of the referent. The item(s) picked out by
193 the color adjective could be located either on the left (and middle), or on the right (and middle).
194 In contrast to Experiments 2 and 3, this allowed us to reduce the number of levels on the factor
195 location from 3 (left, middle and right) to two (left and right), which allowed us to collect more
196 trial per condition while keeping the overall experiment duration low, which was a concern after
197 pilot subjects indicated that the task was tiring. Incidentally, this simplification also allowed us
198 to make the factor location meaningful in both resolving and non-resolving condition, i.e., even
199 when the adjective did not resolve reference its meaning singled out one side of the screen (left or
200 right) over the other. The subsequent noun resolved reference to one of the the two items singled
201 out by the adjective with equal probability. This ensured that participants could not predict the
202 referent of the noun after hearing a non-resolving adjective. Together with the fact that questions
203 were about relationships to another item on the display, this assured that participants had to
204 pay attention to the whole display rather than being able to strategically attend to only some
205 information.

206 For the analysis of adjectives, this resulted in a 2 (reference resolution: resolving or non-
207 resolving) $\times$ 2 (referent location: left or right) design. In each cell of this 2×2 design targets were
208 constructed by permuting the 7 colors with the 7 shapes, resulting in 49 trials per condition and
209 196 trials in total. Our analysis thus compared linguistic expressions that were identical between
210 conditions, with the conditions differing only in the relationship of the expression to the prior
211 visual world context. The remaining aspects of the designs were pseudo-randomized but balanced
212 between conditions, and the order of trials was randomized for each participant. On trials in which
213 the noun resolved reference, the referent could also be located in the middle of the display (late

214 referents left: 33; middle: 28; right: 37).

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

226 2.1.3 Procedure

227 Before the MEG recording, participants’ head shape was digitized using an optical scanner (Pol-
228 hemus FastSCAN, www.polhemus.com). Five fiducial points were marked in the digital scans and
229 coils were attached to those points on participants’ face. The spatial position of those coils with
230 respect to the MEG sensors was recorded before and after each MEG recording session, and were
231 later used to determine the position of the participant’s head with respect to the MEG sensors.
232 The head shape scans were then used to position a virtual head and brain model with respect to
233 the sensor position, in order to constrain the source localization of the MEG data.

234 Participants were given instructions on the reference task before entering the MEG acquisition
235 chamber. They were allowed to perform as many practice trials as they wished until they felt
236 comfortable doing the task (practice trials used stimuli that did not occur in the main experiment).
237 Inside the magnetically shielded MEG acquisition chamber, participants lay in a supine position
238 and stimuli were projected onto a screen at comfortable viewing distance above their eyes. The
239 experimenters communicated with participants via a microphone and headphones. Participants

240 were instructed to blink as little as possible during the presentation of the stimuli. They were told
241 that if they needed a break they could withhold their yes/no response at the end of a trial until they
242 felt comfortable to continue. In regular intervals throughout the experiment they were informed
243 of the progress in the experiment with a text display and had the opportunity to take a short, self-
244 terminated break. Stimuli were presented with MATLAB using psychtoolbox (psychtoolbox.org)
245 and ptbwrapper (code.google.com/p/ptbwrapper). On average, the task lasted approximately 23
246 minutes from first to last trial.

247 **2.1.4 MEG data acquisition and analysis**

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

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

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

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

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

280 **Statistical analysis** To evaluate the effects of our stimulus manipulation on the localized MEG
281 activity, we employed spatio-temporal cluster tests within a broad search area defined on the basis
282 of the fMRI study of referential ambiguity discussed in the Introduction (Nieuwland, Petersson, &
283 Van Berkum, 2007). For each of the brain regions whose response differentiated between referen-
284 tially ambiguous and unambiguous pronouns we used the peak coordinates provided by Nieuwland
285 and colleagues. Peaks were located in medial parietal, lateral parietal, medial frontal and lateral
286 frontal regions in each hemisphere and a right superior frontal region. For each peak, we selected
287 the closest vertex on the fsaverage white matter surface as seed. Around those seed vertices, the
288 search area was then grown on the white matter surface up to a geodesic distance of 50 mm. We
289 excluded from the final search area midline sources lying in the corpus callosum and subcortical
290 structures according to the PALS-B12 atlas (Van Essen, 2005). The final search area is shown in
291 Figure 2, colored according to the location of the original seed vertices.

Figure 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.

292 Within those regions we analyzed current estimates using spatio-temporal cluster based per-
 293 mutation tests to maintain strong control over type I error (Nichols & Holmes, 2002; Maris &
 294 Oostenveld, 2007). For each relevant statistical comparison, we computed maps of F - or t -values
 295 over space and time. We thresholded these maps at a F - or t -value equivalent to an uncorrected
 296 $p = 0.05$. We then formed clusters of suprathreshold values adjacent in space and time. Adjacency
 297 in space was determined based on neighborhood of sources along the cortical surface. In addition,
 298 because surface connections in the FreeSurfer anatomical model do not cross the midline, each
 299 source vertex along the midline was connected to the closest vertex in the opposite hemisphere
 300 if that vertex was less than 15 mm away. This was done to better capture medial clusters of
 301 activation, which would be expected to involve sources in both hemispheres due to spatial disper-
 302 sion of MEG source localization (see Hauk et al., 2011). As critical cluster statistic we used the
 303 exceedance mass, i.e. the sum of all t - or F -values in the cluster, which has the advantage of being
 304 sensitive to extension as well as magnitude of the effect (Nichols & Holmes, 2002). We computed
 305 distributions for this statistic by collecting the maximal statistic in 10,000 random permutations
 306 of the original data, shuffling the condition labels of the condition averages within subjects. In
 307 each permutation, one maximum cluster statistics was collected across the whole search region to
 308 correct the resulting p -values for multiple comparison across the whole search region. For each
 309 cluster in the original comparison we thus calculate a p -value as the proportion of the permutation
 310 distribution above the cluster’s own exceedance mass.

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

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

321 In order to maximize the power of our analysis we did not introduce an exclusion criterion based
322 on performance. This was also justified because incorrect answers could have many reasons, only
323 one of which was failure to resolve the question-initial referential expression. Indeed, identification
324 of the referent was one of the earlier cognitive operations required in answering the questions. The
325 probably more difficult aspect of the task was to perform the mental comparison of the relational
326 statement, indicating which shape was *next to* or *opposite* the referent, with the display. Many
327 incorrect answers were probably due to failure in this task despite successful identification of the
328 referent, and the trials would thus still be suited to analyze the neural correlate of the resolution
329 of the initial referential expression.

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

337 2.2 Results and discussion

338 2.2.1 Behavioral results

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

343 2.2.2 MEG results

344 The analysis with a 2 (reference resolution) by 2 (location: left vs right) ANOVA yielded two
345 significant clusters with a main effect of reference in the medial parietal lobe region, characterized
346 by an increase in activity when reference could be resolved (255-325 ms, $p = .036$ and 435-500,
347 $p = .046$). Figure 3 (top panel) illustrates the spatial extent of the clusters along with the time
348 course of activation in the region defined by their extent. Both clusters included sources in both
349 hemispheres. While it is possible that the effect was caused by bilateral medial parietal activation,
350 the medial parietal sources of the two hemispheres lie in close proximity, often less than 15 mm
351 apart; in view of the spatial dispersion of MEG source estimates it would be premature to draw
352 any conclusions about the lateralization of the effect, and a true source in either hemisphere could
353 result in an effect in both hemispheres (see Hauk et al., 2011). Figure 4 illustrates differences
354 between reference resolving and non-resolving adjectives across the whole brain.

355 The design of Experiment 1 was symmetric in that whenever the adjective did not resolve
356 reference, the subsequent noun did. Consequently, we could also compare reference-resolving nouns
357 with non-resolving nouns (the latter being cases where reference had already been resolved by the
358 adjective). Including the location factor in the design would have lead to an unbalanced model
359 with an empty cell, because in trials with resolving adjectives, references had always been resolved
360 towards the left or right item, whereas nouns also resolved reference to the middle with equal
361 probability as to either side. For this reason we collapsed across location condition for the purpose
362 of analyzing the nouns. Based on the location and timing of the two clusters in the response to the

Figure 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 & Masson, 1994).

Figure 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.

363 adjective we performed a one-tailed t -test in the medial parietal region in the time window from
 364 250 - 500 ms (Figure 3, bottom panel). This analysis revealed one significant cluster (340-445 ms,
 365 $p = .042$).

366 In sum, the combination of our adjective and noun analyses yielded a consistent effect pattern
 367 in medial parietal cortex, with an increase in activity whenever reference could be resolved.

368 Before these results can be attributed to reference resolution, however, it is necessary to dis-
 369 entangle potentially covarying processes. Specifically, reference resolving adjectives also enabled
 370 immediate pre-activation of the subsequent noun: matching the word *blue* to a picture of a blue
 371 heart forms a strong prediction that the noun will be *heart*. If such predictive processes take place
 372 on the adjective, one would expect less corresponding neural activity on the subsequent noun,

373 resulting in exactly the pattern that we observe in our data: more activity on the resolving and
374 predicting adjectives followed by less activity on the non-resolving, pre-accessed nouns. Neurobi-
375 ologically, this hypothesis is at least plausible, given that medial parietal cortex has been argued
376 to be part of a “semantic network” (Binder, Desai, Graves, & Conant, 2009). To address this
377 issue, we conducted two follow-up experiments to assess the extent to which the effects obtained
378 in Experiment 1 might be explainable in terms of lexical prediction of the post-adjectival noun. In
379 Experiment 2 this was achieved by a full crossing of the factors reference resolution and prediction
380 and in Experiment 3, we conducted an Arabic version of Experiment 1, where, due to a different
381 word order, reference resolution did not covary with lexical prediction.

382 **3 Experiment 2**

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

397 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 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 Figure 5). We selected colors and shapes for maximal distinctness in the visual displays, and with the requirement that they completely filled out the area of the aperture that covered the shape on some trials. For broader validity we also

423 included the object in the middle of the visual world displays as a referent in the adjective resolving
 424 condition. Within each cell of the 2 (reference) $\times$ 2 (prediction) $\times$ 3 (location) design we included
 425 all possible referents based on permutation of colors and shapes, i.e. 30. Altogether this resulted
 426 in 360 trials.

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

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

435 (5) Was the pink fish next to a boat?

436 (6) Was the blue heart with stripes in the middle?

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

440 3.1.3 MEG data acquisition and analysis

441 MEG data were collected with a 208-channel whole-head axial gradiometer system (Kanazawa
 442 Institute of Technology, Nonoichi, Japan) sampling at 1000 Hz with a 200 Hz low-pass filter and

443 a 50 Hz notch filter. Raw data were noise-reduced with the continuously adjusted least squares
 444 method method (Adachi et al., 2001, see Experiment 1 methods) and low pass filtered at 40 Hz.
 445 The data for Experiment 2 were acquired at the NYU Abu Dhabi site. Due to its remote location
 446 this facility exhibits lower levels of environmental magnetic noise than the NYU New York site.
 447 Consequently, no high pass filter had to be applied (although the electronics of the machine lead to
 448 an attenuation of low frequencies equivalent to a 0.16 Hz high pass filter). We did not have access to
 449 structural MRIs for any of the participants, and used the fsaverage brain for all participants. Apart
 450 from those differences, data analysis was analogous to Experiment 1. We analyzed source localized
 451 MEG responses to the adjectives with a 2 (resolving vs not resolving reference) by 2 (predicting
 452 vs not predicting the noun) by 3 (location of the referent, left, middle or right) ANOVA in the
 453 time window spanning both clusters found in Experiment 1 (250-500 ms). Based on the findings
 454 of Experiment 1, we restricted our primary search for clusters to the left and right medial parietal
 455 regions (i.e., we calculated p -values for clusters that were corrected for multiple comparisons across
 456 those two regions). In a second, more exploratory stage we included all regions from Experiment
 457 1.

458 **3.1.4 EEG data acquisition and analysis**

459 Data were acquired concurrently to the MEG recordings from 31 EEG and 3 EOG electrodes
 460 attached to an elastic cap at standard positions in the international 10-20 system (Easycap, Ger-
 461 many) at a sampling rate of 1000 Hz. Impedances were kept below 10 k Ω .

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

467 We applied the same cluster permutation method that we also used for MEG source data to
 468 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 Figure 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 correct 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. Figure 7 illustrates differences between reference resolving and non-resolving adjectives across the whole brain.

Main Effect of Reference Resolution

Main Effect of Predictiveness

Figure 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 Figure 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, Morgan-Short, & Luck, 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

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

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

504 The MEG results from Experiment 2 confirm a medial parietal involvement in successful ref-
 505 erence resolution as well as suggesting a possible involvement in predictive processing. This ob-
 506 servation begs the question whether the reference- and prediction related effects are two different
 507 properties of the same response, or whether they reflect two different responses. The former could
 508 be explained by a response that increases as more information about the referent becomes avail-
 509 able. The latter possibility is somewhat favored by the spatial separation between the different
 510 clusters, as visual inspection of figure 6 suggests that the prediction-related response is somewhat
 511 more anterior and the reference-related response somewhat more posterior.

512 In contrast to Experiment 1, we did not find early effects around 260 ms after adjective onset

Figure 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.

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, $p = .037$). Figure 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

528 somewhat shorter than the MEG cluster is most likely due to the specific subgroup of subjects
529 involved: When we repeated the analysis of the MEG source data in this subgroup included in
530 the EEG analysis in the 390-460 ms time window, the medial parietal cluster with main effect of
531 reference also lasted from 390 to 420 ms.

532 Our result suggests that reference resolution in visual displays can also be tracked with EEG
533 and manifests in a negative going component with maximum at posterior electrodes. Prior EEG
534 studies of referential processing commonly assume that a more positive signal in the the N400 time
535 window reflects contextual facilitation (e.g. Burkhardt, 2006). It is possible that our participants
536 expected the color adjective associated with two items on the display more than the adjective
537 associated with just one item. However, in the context of the experiment both adjectives were
538 equiprobable. Hence our results suggest the possibility that a potential difference in this time
539 window could also reflect referential processing.

540 4 Experiment 3

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

547 In Arabic, adjectives follow nouns rather than preceding them. Consequently, the reference-
548 resolving *al-jml* (“the-camel”, ALA-LC transliteration) can be contrasted with the visually identi-
549 cal, but non-resolving *al-jml* in *al-jml al-‘āzrq* (“the-camel the-blue”), where reference is resolved
550 on the color adjective (see Figure 10). In this design, referential expressions only contained a color
551 adjective when the noun alone was referentially ambiguous, i.e. questions in the noun-resolving
552 condition were as simple as *Hal al-jml ‘lá al-ysār?* – “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 Figure 9). As in the previous experiments, each possible color/shape combination occurred once within each cell of the Reference $\times$ Location (left/middle/right) design, for a total of 288 trials.

Figure 9: Experiment 3 Stimuli: Colors and shapes used in Experiment 3

Questions were constructed such that the word following the referential expression was not

571 predictable. The continuation was one of six words (*left, right, on, next-to, between* or *in*), each
572 with equal probability independent of the location of the referent. Questions probed absolute and
573 relative locations, as in examples 7 and 8, respectively:

هل الزهرة في الوسط؟
the-middle in the-flower question-particle

574 (7) Was the flower in the middle?

هل النجمة الخضراء بجانب أرنب؟
rabbit beside the-green the-star question-particle

575 (8) Was the green star next to a rabbit?

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

579 MEG data acquisition and analysis were analogous to Experiment 2. A major difference be-
580 tween English and Arabic writing is that the Arabic article is fused with the noun. Consequently,
581 upon perceiving the target word readers in Experiment 3 had to process the direct article, whereas
582 readers in our English studies had already had time to process it since the article had been pre-
583 sented separately 400 ms before the target word. To account for this difference and allow additional
584 time for processing the direct article, we extended the time window for analysis by 100 ms to to
585 250-600 ms. Based on the findings from the previous experiments the analysis was constrained
586 to the medial parietal region. A 2 (reference) by 3 (location, left/middle/right) ANOVA did not
587 yield significant clusters, but the previous experiments justify applying a one tailed test looking
588 for clusters with an increase in activation for reference resolving over non-resolving nouns.

589 4.2 Results and discussion

590 4.2.1 Behavioral results

591 Average performance was 87.9% correct answers, with individual participants ranging from 70.8%
592 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 Figure 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. Figure 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 that the effect is not dependent on a specific age composition.

Figure 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 left medial parietal region revealed a reference $\times$ location interaction, indicating an increase in activation when the noun resolved reference towards the left or the right side of the display, but not when it resolved reference towards the middle. The analysis in the lateral parietal region revealed a main effect of reference in two clusters, again indicating an increase in activation when reference was resolved. Details on the cluster plots are analogous to Figure 3.

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

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

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,

634 whereas readers of Arabic had to process the definite article upon presentation of the target word.
635 A possible explanation for the early response in Experiment 1 is the predictability of the stimulus
636 material. In Experiment 1, all questions were constructed according to the same schema by filling
637 in a color, two shapes and a relation (*opposite* or *beside*; see examples 1 and 2) and referents of
638 reference resolving adjectives could only be in two possible locations, on the left and the right
639 side of the display. This might have lead readers to engage in strategic processing which allowed
640 them to extract referential or predictive information more quickly than in Experiments 2 and 3,
641 where stimuli were more varied. Overall, the timing of our results is consistent with EEG studies
642 of referential ambiguity which place the time point at which referential ambiguity is registered in
643 the 300-400 ms range (e.g. Van Berkum et al., 1999; Nieuwland & Van Berkum, 2006).

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

653 An open question is what level of representation the medial parietal response is related to, a
654 question that is tied to the question on the role of medial parietal lobe in cognition in general.
655 Based on our results there is a continuum of possibilities ranging from a visual short term buffer
656 to more abstract, discourse level representations. On the one hand, our participants needed a
657 representation of the referential domain, the visual display, which we assume to reside in visual
658 short term memory. Reference resolution amounted to identifying and selecting an entity in that
659 representation. An increase in activation could thus reflect reactivation of the representation of
660 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 & Trimble, 2006) as well as retrosplenial cortex (Vann, Aggleton, & Maguire, 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, Pelle, Berthoz, & Committeri, 2010) as well as when answering questions about spatial relations in visually presented (Wallentin, Roepstorff, Glover, & Burgess, 2006) or described (Wallentin, Weed, Østergaard, Mouridsen, & Roepstorff, 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 & Tian, 2006) and spatial attention (e.g. Corbetta & Shulman, 2011). It is thus plausible that activity in this region reflects coding of the spatial location of the referent or with an attentional shift to its position, and might even be involved in planning of a saccade towards the referent. Notably, 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.

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

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

713 Together with results cited in the Introduction, our results suggest a cortical division of labor
714 during reference resolution. While our results associate parietal cortex with tracking referents on

715 a basic level, in line with the results of Boiteau et al. (2014) and Almor et al. (2007), frontal
716 lobe activity seems to be associated with higher order decision making processes in non-trivial
717 referential situations such as referential ambiguity (Nieuwland, Petersson, & Van Berkum, 2007;
718 McMillan et al., 2012).

719 An important distinction that might warrant future attention is the mechanism of reference
720 resolution and its relationship to processing of the sets of referential candidates. One way to model
721 reference resolution is through sets: Referential processing starts with a set of all entities in the
722 referential domain, and any information is used to narrow down this set. For example, if the
723 current referential candidates is a set of yellow and blue objects, the adjective *blue* constrains the
724 set of referential candidates to the subset of blue objects. The cognitive act of reference resolution
725 can be seen as pruning a set of potential referents until a unique object is left. In the context of our
726 paradigm, the set of potential referents is initially the set of the three objects in the display, which
727 is then pruned with content words like color adjectives and shape-describing nouns until a single
728 item remains in the set, the referent. In terms of this model, the main contrasts in Experiments
729 1-3 compared reduction of the candidates from 3 to 1 with a reduction from 3 to 2 (e.g., *blue* in
730 the context with one vs two blue objects). The analysis of the noun in Experiment 1 compared
731 reduction from 2 to 1 with no reduction at all (when the preceding adjective had already narrowed
732 the set to 1). Consequently, the medial parietal effect we obtained could also be related to a brain
733 response that scales with the number of referents that could be excluded (excluding 2 vs 1, or 1
734 vs 0). Future work should address this important issue.

735 A further issue concerns the properties of the referential domain. Visual world studies are
736 sometimes criticized for involving artificially small referential domains. This criticism could be
737 leveled against our results in particular. While some real world referential domains are certainly
738 much larger, there are also situations where speakers and listeners refer to items from a constrained
739 set. A good example that comes close to the kinds of domains we studied might be deliberations
740 about articles of clothing, as when someone needs help deciding between *the green sweater*, *the*
741 *green shirt* or *the white shirt*. Such situations might or might not recruit the same cognitive

742 mechanism as more out-of-the-blue referential situations with larger domains. In particular, the
743 decision to accept a specific item as the referent might be simplified when the given property (e.g. a
744 color adjective) only has to be compared with three items from short term memory as opposed to a
745 larger domain from long term memory. As most research does, we thus identified a specific example
746 of the phenomenon we intend to study, one that lends itself to a particular operationalization, and
747 future research will have to show whether our results generalize to different scenarios, or whether
748 different cognitive processes are involved in referential processing depending on the nature of the
749 referential situation.

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

762 It is in order to point out some methodological limitations. While visual world studies typically
763 employ spoken language stimuli, emphasizing naturalness of the stimulus material, we opted for
764 visual stimuli. This is in line with a long tradition in MEG/EEG research taking advantage of the
765 more highly controllable nature of visual language stimuli. Important for MEG/EEG, the temporal
766 characteristics of written stimuli, presented word by word, can be strictly controlled, whereas
767 natural speech stimuli introduce more temporal variability into the signal. As a consequence of
768 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 & 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 to written (Van Berkum et al., 1999) as well as spoken material (Nieuwland, Otten, & Van Berkum, 2007).

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 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 suggests that an attentional shift towards the location of the referent slightly precedes a medial

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

800 7 Acknowledgments

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

804 References

- 805 Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., & Ochiai, M. (2001). Reduction of non-
806 periodic environmental magnetic noise in MEG measurement by continuously adjusted least
807 squares method. *IEEE Transactions on Applied Superconductivity*, 11(1), 669–672. doi:10.
808 1109/77.919433
- 809 Allen, T. A. & Fortin, N. J. (2013). The evolution of episodic memory. *Proceedings of the National*
810 *Academy of Sciences*, 110, 10379–10386. doi:10.1073/pnas.1301199110
- 811 Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of
812 Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models.
813 *Journal of Memory and Language*, 38(4), 419–439. doi:10.1006/jmla.1997.2558
- 814 Almor, A., Smith, D. V., Bonilha, L., Fridriksson, J., & Rorden, C. (2007). What is in a name?
815 Spatial brain circuits are used to track discourse references. *Neuroreport*, 18(12), 1215–1219.
816 doi:10.1097/WNR.0b013e32810f2e11

817 Altmann, G. T. M. & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the
818 domain of subsequent reference. *Cognition*, 73(3), 247–264. doi:10.1016/S0010-0277(99)
819 00059-1

820 Altmann, G. T. M. & Kamide, Y. (2007). The real-time mediation of visual attention by lan-
821 guage and world knowledge: Linking anticipatory (and other) eye movements to linguistic
822 processing. *Journal of Memory and Language*. Language-Vision Interaction, 57(4), 502–518.
823 doi:10.1016/j.jml.2006.12.004

824 Anderson, J. E. & Holcomb, P. J. [Phillip J.]. (2005). An electrophysiological investigation of the
825 effects of coreference on word repetition and synonymy. *Brain and Language*, 94(2), 200–216.
826 doi:10.1016/j.bandl.2005.01.001

827 Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic
828 System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies.
829 *Cerebral Cortex*, 19(12), 2767–2796. doi:10.1093/cercor/bhp055

830 Boiteau, T. W., Bowers, E., Nair, V. A., & Almor, A. (2014). The neural representation of plural
831 discourse entities. *Brain and Language*, 137, 130–141. doi:10.1016/j.bandl.2014.08.003

832 Boudewyn, M. A., Long, D. L., Traxler, M. J., Lesh, T. A., Dave, S., Mangun, G. R., ... Swaab,
833 T. Y. (2015). Sensitivity to Referential Ambiguity in Discourse: The Role of Attention,
834 Working Memory, and Verbal Ability. *Journal of Cognitive Neuroscience*, 27(12), 2309–
835 2323.

836 Brodbeck, C., Gwilliams, L., & Pykkänen, L. (2015). EEG can track the time course of successful
837 reference resolution in small visual worlds. *Frontiers in psychology*, 6, 1787. doi:10.3389/
838 fpsyg.2015.01787

839 Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence
840 from event-related brain potentials. *Brain Lang*, 98, 159–168. doi:10.1016/j.bandl.2006.04.005

841 Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory: *NeuroReport*,
842 18(17), 1851–1854. doi:10.1097/WNR.0b013e3282f1a999

843 Cavanna, A. E. & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and
844 behavioural correlates. *Brain*, 129(3), 564–583. doi:10.1093/brain/awl004

845 Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Cir-
846 cumscribing Referential Domains during Real-Time Language Comprehension. *Journal of*
847 *Memory and Language*, 47(1), 30–49. doi:10.1006/jmla.2001.2832

848 Committeri, G., Galati, G., Paradis, A.-L., Pizzamiglio, L., Berthoz, A., & LeBihan, D. (2004). Ref-
849 erence Frames for Spatial Cognition: Different Brain Areas are Involved in Viewer-, Object-,
850 and Landmark-Centered Judgments About Object Location. *Journal of Cognitive Neuro-*
851 *science*, 16(9), 1517–1535. doi:10.1162/0898929042568550

852 Corbetta, M. & Shulman, G. L. (2011). Spatial Neglect and Attention Networks. *Annual Review*
853 *of Neuroscience*, 34(1), 569–599. doi:10.1146/annurev-neuro-061010-113731

854 Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren,
855 E. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-
856 resolution imaging of cortical activity. *Neuron*, 26(1), 55–67. doi:10.1016/S0896-6273(00)
857 81138-1

858 Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye move-
859 ments as a window into real-time spoken language comprehension in natural contexts. *Journal*
860 *of Psycholinguistic Research*, 24(6), 409–436. doi:10.1007/BF02143160

861 Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language
862 network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain*
863 *Mapping*, 29(5), 581–593. doi:10.1002/hbm.20422

864 Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-resolution intersubject
865 averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4),
866 272–284. doi:10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4

867 Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function.
868 *Physiological Reviews*, 91(4), 1357–1392. doi:10.1152/physrev.00006.2011

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. doi:10.1007/s00221-010-2168-8

Grainger, J. & Holcomb, P. J. [Phillip 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. doi:10.1111/j.1749-818X.2008.00121.x

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., . . . Hämäläinen, M. S. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 267. doi:10.3389/fnins.2013.00267

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., . . . Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, 86, 446–460. doi:10.1016/j.neuroimage.2013.10.027

Hagoort, P. & Indefrey, P. (2014). The Neurobiology of Language Beyond Single Words. *Annual Review of Neuroscience*, 37(1), 347–362. doi:10.1146/annurev-neuro-071013-013847

Hämäläinen, M. S. & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & Biological Engineering & Computing*, 32(1), 35–42. doi:10.1007/BF02512476

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. doi:10.1016/j.neuroimage.2010.09.053

Huetig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*. Visual search and visual world: Interactions among visual attention, language, and working memory, 137(2), 151–171. doi:10.1016/j.actpsy.2010.11.003

Hutton, S. B. (2008). Cognitive control of saccadic eye movements. *Brain and Cognition*. A Hundred Years of Eye Movement Research in Psychiatry, 68(3), 327–340. doi:10.1016/j.bandc.2008.08.021

- 896 Kahn, I., Andrews-Hanna, J. R., Vincent, J. L., Snyder, A. Z., & Buckner, R. L. (2008). Distinct
897 Cortical Anatomy Linked to Subregions of the Medial Temporal Lobe Revealed by Intrinsic
898 Functional Connectivity. *Journal of Neurophysiology*, *100*(1), 129–139. doi:10.1152/jn.00077.
899 2008
- 900 Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incre-
901 mental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory*
902 *and Language*, *49*(1), 133–156. doi:10.1016/S0749-596X(03)00023-8
- 903 Kintsch, W. (1988). The Role of Knowledge in Discourse Comprehension - a Construction Inte-
904 gration Model. *Psychol Rev*, *95*, 163–182. doi:10.1037/0033-295x.95.2.163
- 905 Lambrecht, K. (1994). *Information structure and sentence form: Topics, focus, and the mental*
906 *representations of discourse referents*. Cambridge studies in linguistics. New York, NY, US:
907 Cambridge University Press.
- 908 Ledoux, K., Gordon, P. C., Camblin, C. C., & Swaab, T. Y. (2007). Coreference and lexical rep-
909 etition: Mechanisms of discourse integration. *Memory & Cognition*, *35*(4), 801–815. doi:10.
910 3758/BF03193316
- 911 Loftus, G. R. & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs.
912 *Psychonomic Bulletin & Review*, *1*(4), 476–490. doi:10.3758/BF03210951
- 913 Luck, S. J. & Vogel, E. K. (1997). The capacity of visual working memory for features and con-
914 junctions. *Nature*, *390*(6657), 279–281. doi:10.1038/36846
- 915 Lynch, J. & Tian, J.-R. (2006). Cortico-cortical networks and cortico-subcortical loops for the
916 higher control of eye movements. In *Progress in Brain Research* (Vol. 151, pp. 461–501).
917 Elsevier. doi:10.1016/S0079-6123(05)51015-X
- 918 Maguire, E. A. (2001). The retrosplenial contribution to human navigation: a review of lesion and
919 neuroimaging findings. *Scandinavian journal of psychology*, *42*(3), 225–238. doi:10.1111/
920 1467-9450.00233
- 921 Maris, E. & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data.
922 *Journal of Neuroscience Methods*, *164*(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024

923 McMillan, C. T., Clark, R., Gunawardena, D., Ryant, N., & Grossman, M. (2012). fMRI evidence
 924 for strategic decision-making during resolution of pronoun reference. *Neuropsychologia*, *50*(5),
 925 674–687. doi:10.1016/j.neuropsychologia.2012.01.004

926 McNorgan, C. (2012). A meta-analytic review of multisensory imagery identifies the neural corre-
 927 lates of modality-specific and modality-general imagery. *Frontiers in Human Neuroscience*,
 928 *6*. doi:10.3389/fnhum.2012.00285

929 Nichols, T. E. & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimag-
 930 ing: A primer with examples. *Human Brain Mapping*, *15*(1), 1–25. doi:10.1002/hbm.1058

931 Nieuwland, M. S., Otten, M., & Van Berkum, J. J. A. (2007). Who are You Talking About?
 932 Tracking Discourse-level Referential Processing with Event-related Brain Potentials. *Journal*
 933 *of Cognitive Neuroscience*, *19*(2), 228–236. doi:10.1162/jocn.2007.19.2.228

934 Nieuwland, M. S., Petersson, K. M., & Van Berkum, J. J. A. (2007). On sense and reference:
 935 examining the functional neuroanatomy of referential processing. *NeuroImage*, *37*(3), 993–
 936 1004. doi:10.1016/j.neuroimage.2007.05.048

937 Nieuwland, M. S. & Van Berkum, J. J. A. (2006). Individual differences and contextual bias in
 938 pronoun resolution: Evidence from ERPs. *Brain Research*, *1118*(1), 155–167. doi:10.1016/j.
 939 brainres.2006.08.022

940 Nieuwland, M. S. & Van Berkum, J. J. A. (2008). The Neurocognition of Referential Ambiguity in
 941 Language Comprehension. *Language and Linguistics Compass*, *2*(4), 603–630. doi:10.1111/
 942 j.1749-818X.2008.00070.x

943 Ranganath, C. & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature*
 944 *Reviews Neuroscience*, *13*(10), 713–726. doi:10.1038/nrn3338

945 Sedivy, J. C., K. Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental
 946 semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147. doi:10.
 947 1016/S0010-0277(99)00025-6

- 948 Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and
 949 spoken language comprehension: Effects of visual context on syntactic ambiguity resolution.
 950 *Cognitive Psychology*, 45(4), 447–481. doi:10.1016/S0010-0285(02)00503-0
- 951 Tanenhaus, M. K. & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philo-*
 952 *sophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1105–1122.
 953 doi:10.1098/rstb.2007.2162
- 954 Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration
 955 of visual and linguistic information in spoken language comprehension. *Science*, 268(5217),
 956 1632–1634. doi:10.1126/science.7777863
- 957 Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can
 958 produce artifactual effects and incorrect conclusions in ERP studies of language and cog-
 959 nition: High-pass filtering and artifactual ERP effects. *Psychophysiology*, 52(8), 997–1009.
 960 doi:10.1111/psyp.12437
- 961 Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early Referential Context Effects in
 962 Sentence Processing: Evidence from Event-Related Brain Potentials. *Journal of Memory and*
 963 *Language*, 41(2), 147–182. doi:10.1006/jmla.1999.2641
- 964 Van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitterlood, P. (2003). Event-related brain
 965 potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psy-*
 966 *chophysiology*, 40(2), 235–248. doi:10.1111/1469-8986.00025
- 967 Van Essen, D. C. (2005). A Population-Average, Landmark- and Surface-based (PALS) atlas of
 968 human cerebral cortex. *NeuroImage*, 28(3), 635–662. doi:10.1016/j.neuroimage.2005.06.058
- 969 Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the
 970 Word Repetition Effect with Event-Related Potentials. *Journal of Cognitive Neuroscience*,
 971 3(2), 131–150. doi:10.1162/jocn.1991.3.2.131
- 972 van Deemter, K. (2016). *Computational models of referring: a study in cognitive science*. Cam-
 973 bridge, MA: The MIT Press.

- 974 Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do?
975 *Nature Reviews Neuroscience*, 10(11), 792–802. doi:10.1038/nrn2733
- 976 Wallentin, M., Roepstorff, A., Glover, R., & Burgess, N. (2006). Parallel memory systems for talking
977 about location and age in precuneus, caudate and Broca’s region. *NeuroImage*, 32(4), 1850–
978 1864. doi:10.1016/j.neuroimage.2006.05.002
- 979 Wallentin, M., Weed, E., Østergaard, L., Mouridsen, K., & Roepstorff, A. (2008). Accessing the
980 mental space—Spatial working memory processes for language and vision overlap in pre-
981 cuneus. *Human Brain Mapping*, 29(5), 524–532. doi:10.1002/hbm.20413
- 982 Zelinsky, G. J. & Murphy, G. L. (2000). Synchronizing Visual and Language Processing: An
983 Effect of Object Name Length on Eye Movements. *Psychological Science*, 11(2), 125–131.
984 doi:10.1111/1467-9280.00227