Brain and Cognition 77 (2011) 304–314

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Over-specified referring expressions impair comprehension: An ERP study

Paul E. Engelhardt [a,*], Ş. Barış Demiral [b], Fernanda Ferreira [c]

[a] Department of Psychology, Northumbria University, Northumberland Building, Newcastle upon Tyne NE1 8ST, UK
[b] Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
[c] Department of Psychology, University of South Carolina, Barnwell College, Columbia SC 29208, USA

ARTICLE INFO

Article history:
Accepted 14 July 2011
Available online 15 August 2011

Keywords:
Maxim of Quantity
Over-description
Pre-nominal modifier
Adjective
N400
Semantic integration

ABSTRACT

Speakers often include extra information when producing referring expressions, which is inconsistent with the Maxim of Quantity (Grice, 1975). In this study, we investigated how comprehension is affected by unnecessary information. The literature is mixed: some studies have found that extra information facilitates comprehension and others reported impairments. We used an attentional-cueing paradigm to assess how quickly participants could orient attention to an object upon hearing a referring expression, such as *the red square*. If there are two squares differing in color, then the modifier is required. However, if there is only one (red) square, then the modifier is unnecessary. We also recorded event-related potentials (ERPs) in order to investigate online processing. Reaction times were significantly longer for referring expressions that contained extra information, and ERPs revealed a centroparietal negativity (N400) that emerged approximately 200–300 ms after modifier onset. We conclude that referring expressions with an unnecessary pre-nominal modifier impair comprehension performance.

1. Introduction

The Maxim of Quantity contains two parts (Grice, 1975). The first is that speakers should include enough information for an object to be identified. The second is that speakers should not be more informative than necessary. Adult speakers are typically very good at adhering to the first part of the Quantity Maxim. That is, they rarely produce ambiguous or under-specified referring expressions. However, there are numerous language production studies, which have shown that adult speakers will often provide extra or unnecessary information (Belke, 2006; Deutsch & Pechmann, 1982; Engelhardt, Bailey, & Ferreira, 2006; Pechmann, 1989; Sonnenschein, 1984). If an expression, such as *the book,* would be sufficient for a referent to be identified, people have a tendency to produce expressions, such as *the red book* or *the book that's on the table*. We refer to these types of referring expressions as over-descriptions. An over-description is a referential expression that has a modifier, but occurs in a context that does not contain two or more objects of the same type (e.g. two or more books). Such expressions can also be thought of as providing extra information than is necessary for identification. Estimates of how often speakers include extra information varies quite widely, from 10% to 60%, depending on the task and type of modifier.

In the past few years, a debate has emerged concerning how over-descriptions affect comprehension. Grice (1989) discussed two possibilities. The first was that extra information might 'simply be a waste of time', in which case, we might expect over-descriptions to have a minimal effect on comprehension. The second was that listeners might think there is 'something communicatively relevant' about the extra information. In this case, over-descriptions would be misleading about the existence of a set of objects that in fact does not exist. This second possibility suggests that over-descriptions might be detrimental to comprehension, if in fact, listeners are mislead by them.

Several studies have argued that additional information is beneficial for comprehension performance (Arts, 2004; Arts, Maes, Noordman, & Jansen, 2011; Davies & Katsos, 2009; Levelt, 1989; Maes, Arts, & Noordman, 2004). In the Arts et al. (2011) study, participants were presented with three successive screens on each trial. On the first screen, participants read a description, such as *the large square gray button,* and then, they pressed a button to view the second screen.[1] The second screen had four objects (i.e. square, triangle, rectangle, circle) arranged in a 2×2 array. Objects varied in size (large or small) and color (gray or white). At the bottom of each object was a number, and participants had to identify the object in the array and encode the corresponding number. Finally, participants pressed a button to view the final screen, which consisted of a judgment task in which participants had to verify, via yes/no decision, the identity of the number that was under the object. The key reaction time measure was the length of time that participants

* Corresponding author.
E-mail address: paul.engelhardt@northumbria.ac.uk (P.E. Engelhardt).

[1] The study was conducted in Dutch, and so the examples are English translations. Word order preferences however, are similar in Dutch and English.

0278-2626/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandc.2011.07.004

viewed the 2 × 2 object array, which was referred to as the identification time.

In this study, the object was uniquely identifiable with shape information alone (e.g. *the square button*). However, when all three attributes (size, color, and shape) were included, participants were 58 ms faster identifying the object compared to the shape-only instruction. The authors argued, similar to Levelt (1989), that the extra information was useful insofar as it helped create a Gestalt for the object, which facilitated search and identification. In the same study, Arts et al. also looked at information concerning spatial location. They found that including extra information about vertical location (i.e. top vs. bottom) also facilitated performance. For example, if participants read *the square button on top,* then they were 95 ms faster compared to a description, such as *the square button.* This finding suggests that extra information, which restricts the search space can also be utilized to facilitate object identification.

Other studies have concluded that additional information is detrimental to comprehension performance (Engelhardt, 2008; Engelhardt et al., 2006; Grodner & Sedivy, 2011; Sedivy, 2003; Sedivy, 2006; Sedivy, 2007). Engelhardt et al. (2006) reported data from an eye tracking study which showed that listeners were slower to execute instructions that contained an unnecessary prepositional phrase modifier. In their study, participants heard, for example, *put the apple in the box* or *put the apple on the towel in the box.* The visual displays consisted of 2 × 2 arrays, and contained, for these examples, an apple on a towel, a pencil, an empty towel, and an empty box. From the onset of the word *box*, which is the disambiguating word, participants were approximately 1 s slower in executing the instruction with the unnecessary modifier (e.g. *on the towel*). The slowdown was primarily due to the fact that the empty towel receives a substantial proportion of fixations because participants get misled into believing that the empty towel is the destination for the apple. In this case, the unnecessary prepositional phrase modifier leads to temporary confusion about where the apple should be placed.

In summary, the evidence concerning the comprehension of over-descriptions has been mixed. Therefore, the purpose of this investigation was to adjudicate between the two competing views on the comprehension of referential expressions that contain extra (or unnecessary) information. Given the differences in tasks that have been used in previous studies, we thought it best to use simple visual displays and simple instructions. We expected over-descriptions to lead to slowdowns in reaction time, similar to what was reported by Engelhardt et al. (2006). We also collected event-related brain potentials (ERPs) because they have been shown to be sensitive to different types of referential, semantic, and syntactic anomalies during language comprehension (for reviews see: Bornkessel-Schlesewsky & Schlesewsky, 2009; Friederici, 2002Hagoort, Brown, & Osterhout, 1999; Kutas, Van Petten, & Kluender, 2006; Osterhout & Holcomb, 1992; Rugg & Coles, 1997; Van Berkum, 2008). If our processing impairment predictions with respect to over-descriptions are confirmed, then we expected ERPs to reveal the time course in which the language processing system registers the presence of an over-description, and we also expected to observe one of the ERP components that have been previously associated with processing difficulties in language comprehension.

1.1. Electrophysiology of language comprehension

There have been several experiments by Van Berkum and colleagues that have focused on referential processing in discourse (Nieuwland & Van Berkum, 2008; Van Berkum, 2008; Van Berkum, Hagoort, & Brown, 1999; Van Berkum, Koornneef, Otten, & Nieuwland, 2007). In the Van Berkum et al. (1999) study, participants read short discourses that introduced either one or two potential referents. For example, if *Bill* and *John* were both introduced into the discourse and later an ambiguous pronoun, such as *he*, was used, then this situation elicited a sustained negative ERP component with a frontal scalp distribution. The component emerged at approximately 280 ms after the referentially ambiguous pronoun, suggesting that the processing system can quite quickly determine whether a referring expression has a unique referent or not. This type of situation is referred to as under-specification, and it occurs when too little information is provided for unique reference. Van Berkum et al. labeled this the negative ERP component the *Nref*, and it has also been elicited from naturalistic spoken language with ambiguous pronouns (Van Berkum, Brown, Hagoort, & Zwitserlood, 2003).

A second language-related ERP component is the N400. This potential has been most often associated with semantic integration problems (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980; Lau, Phillips, & Poeppel, 2008). For example, if a sentence contains an anomalous word, such as *He spread butter and jam over his socks.*, there will be a negative going potential beginning at 200–300 ms after the onset of the word *socks* and peaking at approximately 400 ms. The N400 can also be elicited in an attenuated form from meaningful but unexpected words. Thus, the N400 is widely viewed to result from a combinatorial semantic-integration process in which individual words are integrated with the previous context (Kutas & Federmeier, 2000). A second view on the N400, is that it primarily reflects the difficulty of lexical access from long term memory. By this account, contextual support allows certain words to be accessed and retrieved more easily from long term memory (Federmeier & Kutas, 1999; Van Berkum et al., 1999). In highly constraining contexts, such as sentences with high cloze probability, the comprehension system can make specific predictions about particular lexical items, in which case the comprehension system can get a head start by pre-activating words or conceptual information. The N400 in these "predictable" cases is, therefore, reduced.

Another prominent language-related ERP component is the P600, which is typically elicited from syntactic violations, such as subject-verb agreement. Osterhout and Holcomb (1992) presented participants with ungrammatical sentences (e.g. **The woman persuaded to sell the stock.*) and un-preferred structures (e.g. *The broker persuaded to sell the stock was sent to jail.*). They found a sustained positive shift in the ERP waveform that was largest over central-parietal sites. The amplitude of the P600 was larger for ungrammatical sentences compared to the less preferred reduced relative structures. Later studies have argued that the P600 is not exclusive to the detection of syntactic anomalies, but instead, reflects reanalysis or repair processes after a problem has been encountered (Friederici, Hahne, & Meklinger, 1996). The P600 is sometimes accompanied by an earlier negativity over anterior scalp sites. Both Osterhout and Holcomb (1992) and Friederici et al. (1996) observed this early negativity in response to a word category violation. In a particular syntactic context, certain classes of words (e.g. noun, verb, etc.) are allowed and others are not, and so when an unexpected word occurs in a context where it should not, it elicits an early negativity followed by a P600.

1.2. Current study

The purpose of this study was to investigate the processing of referential expressions that contained extra information, which we refer to as over-descriptions. Previous studies, examining reaction time differences, have concluded that too much information either facilitates (e.g. Arts et al., 2011) or impairs comprehension performance (e.g. Engelhardt et al., 2006). In order to distinguish these competing positions, we constructed a referential processing task, in which participants heard an utterance that contained an adjective (e.g. *look to the big triangle*) and referred to one of two objects. The visual displays consisted of two objects presented side by side on the computer screen. The objects were either of the same type (e.g. two triangles) or different (e.g. a triangle and a square). When the objects are of the same type, a modifier is required to distinguish between them, and when the objects are different, a modifier is unnecessary, which results in an over-description. Reaction times and ERPs were time locked to the onset of the modifier, which in our study was either a size or color adjective.

We predicted that if over-descriptions impair comprehension performance, then we would observe a difference in reaction time when the display contains two objects of the same type compared to when the objects are different. More specifically, we predicted that participants would be slower when the display contains two different objects, because in this situation the modifier is unnecessary (i.e. over-described). For the ERP results, we predicted that if the language comprehension system has difficulty integrating a modifier with a context in which it is not required, then we would observe an ERP component associated with processing difficulty (i.e. N*ref*, N400, or P600). On the other hand, if extra information is beneficial to comprehension, then we would expect faster reaction times when the objects are different compared to when they are the same. In this case, we would also expect little difference in the ERP waveform, and no components suggestive of processing difficulty.

2. Method

2.1. Participants

Fourteen undergraduate students from the University of Edinburgh participated in the experiment. Participants ranged in age from 18 to 26 years ($M = 20.68$, $SD = 2.24$), and 71.0% were female. Participants were native speakers of British English and had normal or corrected-to-normal vision. Each gave informed consent agreeing to participate, and ethical clearance for the study was granted by the University of Edinburgh Faculty Research Committee. Participants were paid £12 for their participation.

2.2. Materials

Four hundred and three visual displays were created: Three were for practice, 160 were experimental items (40 per condition), and 240 were filler trials. For each display, participants heard a single auditory instruction. For the experimental items, 80 of the visual displays had two objects of the same type (see Fig. 1, panels A and C). The instruction for these trials referred to one of the two objects in the display, and each one contained a single pre-nominal adjective modifier. Half of the trials had a size modifier and half had a color modifier. The other 80 experimental trials had two different shapes (see Fig. 1, panels B and D). These displays were also paired with an instruction that had a modifier, and half were size and half were color. These trials were the ones that contained an unnecessary modifier, and as a result, were over-described. For the example displays in Fig. 1, the corresponding instructions were *look to the large star* (panels A and B) and *look to the red star* (panels C and D).

For the filler displays, 160 had two different objects (e.g. star and square). The instructions for these displays did not have a pre-nominal modifier (e.g. *look to the star*), and the objects in these displays were random colors and shapes (see Fig. 2, panel B). The other 80 filler displays had two objects of the same type. These displays were created so that both size and color were manipulated (see Fig. 2, panel A), and the instructions for these trials contained a single pre-nominal modifier. Hence, over the whole experiment there were 320 trials with the "appropriate" amount of modification (i.e. not under-described and not over-described), and 80 trials that contained an extra modifier (i.e. were over-described). Table 1 shows a list of the modifiers and the shapes that were used in the experiment. Objects were 8 cm on either side of the center of the computer screen, and the large objects were 4–4.5 cm and small objects were 2–2.5 cm in size. Images were 600 × 800 pixels.

Two lists were created in which the displays were mirror images of one another. This eliminates any left versus right biases in terms of saliency. The side of the display in which the target was located on was also counterbalanced within each list, meaning that half of the displays had the target on the left and the other half on the right. Furthermore, counterbalancing was also performed

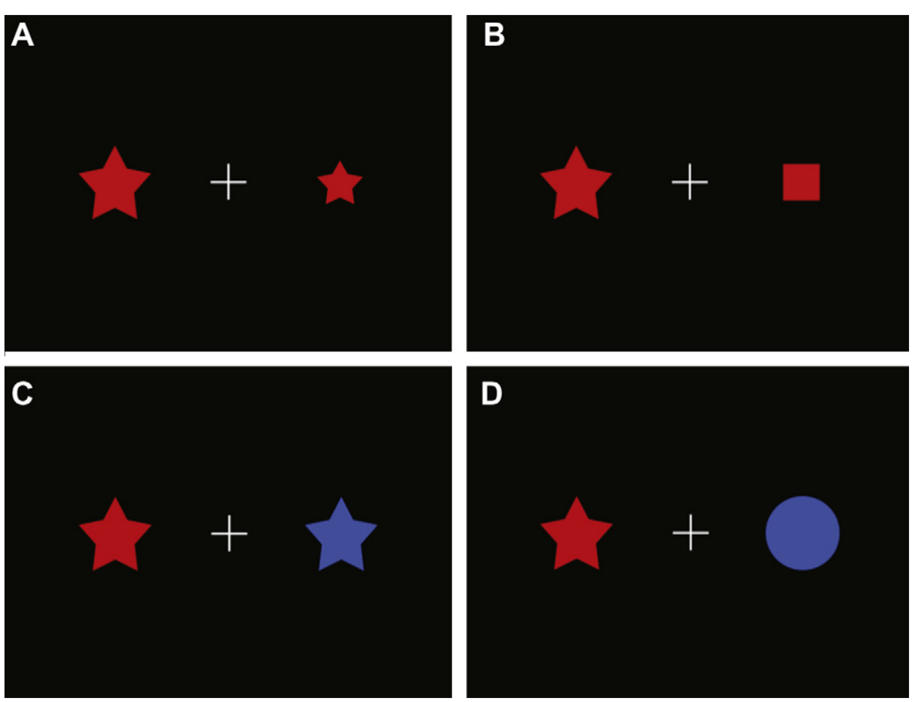

Fig. 1. Sample stimuli. The left panels shows displays with two objects of the same type. The right panels show the displays that have two different objects.

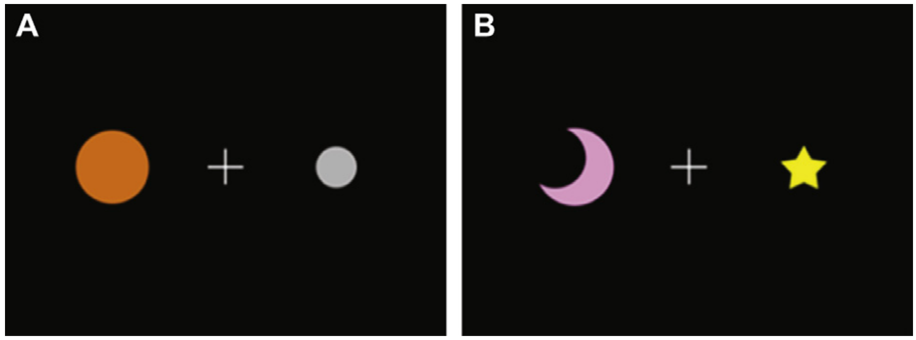

Fig. 2. Sample stimuli for the two types of filler trials in the experiment. Panel A shows a display in which both size and color were manipulated, and the objects were the same shape. Panel B shows a display with two different objects, and this trial was paired with an instruction that did not contain a modifier (e.g. *look to the star*).

regarding the pre-nominal modifier type. In the displays in which the target item was on the left, the modifier was color for half of the trials and size for the other half. The same also held for the displays where the target object was on the right. The onset of the modifier in the spoken instruction was determined by two independent raters. The values for each item were compared, and large differences (i.e. >10 ms) were re-evaluated.

2.3. Apparatus

Electroencephalography (EEG) was recorded with a BioSemi ActiveTwo system. We used an international 10–20 electrode cap configuration with 64 EEG channels. Six other EOG electrodes were placed on the mastoids (as a linked-reference), horizontal cantus of the right (ROC) and left eyes (LOC), and vertically on the top (VEOG) and the bottom of the right eye of the participant. This was done to track blinks and eye movements. Stimulus presentation was programmed using Experiment Builder Software, and the images were presented to participants on a 48 cm (19″) CRT color display monitor. The auditory instructions were played out, via speakers, which were located on either side of the computer monitor. The volume was adjusted so that the instruction could be easily heard, but not uncomfortably loud.

2.3.1. EEG signal processing and data analysis

The initial recording was done at 512 Hz with a BioSemi amplifier, which conducts digital filtering. Low pass filtering is performed in the ADC's decimation filter, which has a 5th order sinc response with a −3 dB point at a 1/5th of the sampling rate (i.e. approximately 100 Hz). Data were then imported to EEGLAB software (Delorme & Makeig, 2004) with the left mastoid as reference. We then, high- and low-pass filtered the data (in that order) with the half-amplitude cutoff values at 0.3 Hz and 40 Hz, respectively. The low-pass filtering used a Finite Impulse Response filter (FIR) with a length of around 150 sampling points. Afterwards, we re-referenced the data to linked mastoids, and down-sampled the data to 256 Hz. Epochs were selected between −300 ms and 1200 ms in which 0 ms was the onset of the critical word (i.e. the adjective). The baseline was a 100 ms time window before the onset of each critical word. After the statistical analyses, we further low-pass filtered the data for visualization purposes when creating individual electrode plots.

Automatic artifact rejection was applied by using eye electrodes ROC, LOC, and VEOG, such that, epochs having activity on these electrodes greater than 70 μV was rejected from the analysis. After rejecting eye-blink artifacts automatically, we did two additional things to ensure that the data was as clean as possible. The first was that we manually eliminated epochs that had a substantial number of artifact containing EEG channels. The second is that we interpolated noisy/bad electrodes with spherical interpolation. This was done in several cases where there were only a few noisy channels. During this manual procedure, we also identified a few epochs with small eye-blinks, which were not detected by the automatic detection process. For all participants, the number of epochs rejected manually was a small fraction (i.e. 2–3%) compared to those that were detected automatically.

Participants with less than 15 trials per condition, after artifact rejection, were eliminated from further ERP and statistical analysis. Table 2 shows a summary of trials per condition that the analyses were based on. Elimination of trials with eye-blinks and eye-movements resulted in the exclusion of five participants. Thus, we report behavioral and electrophysiological results for those remaining fourteen participants, and for the ERP analyses we only included artifact-free trials.

To statistically analyze the ERP data, we created forty 30 ms time windows beginning at the onset of the adjective and continuing to 1200 ms (Demiral, Schlesewsky, & Bornkessel-Schlesewsky, 2008). The spatial arrangement of the 25 regions of interest is shown in Fig. 5. There were eight regions of interest on the left, nine on the mid-line, and eight on the right. We then calculated the areas (via integration) under the ERP activity curves for each condition. We noted all instances where there were at least three consecutive time windows (i.e. 90 ms) that showed the same significant effect for each region of interest. In those instances where there were adjacent time-bins with significant effects, we re-calculated the area under the curve across longer time windows. These longer time windows were then submitted to a statistical analysis, which consisted of 2×2 repeated measures ANOVAs on the averaged data for each region of interest.

2.4. Design and procedure

We used a 2×2 (modifier × display) within subjects design. Modifier was either a size or color adjective, and display refers to whether the visual display contained two objects of the same type or two different objects. Participants completed three practice trials, 160 experimental trials (40 per condition), and 240 filler trials. The filler trials were included in the study so that participants would not see the same type of display and hear the same type of instruction over and over again. Trials were presented in a random order and divided into eight blocks. Each block contained 50 trials. Participants were instructed that they would see an array of two objects and hear an instruction referring to one of the objects. Their task was to indicate which side of the display the target object was on as quickly and accurately as possible by pressing the appropriate button on the keyboard ("A" for left and "L" for right). Participants used their left index finger to press the "A" button and their right index finger to press the "L" button.

Table 1
Sizes, colors, and shapes for the objects used in the experiment.

Words tested	SpFr	WrFr	Concrt	Image	#Lett	#Phon	#Syll
Size adjectives							
Little	700	831	378	502	6	4	1
Small	216	542	402	447	5	4	1
Large	116	361	–	449	5	3	1
Big	549	360	–	463	3	3	1
Means	395	524	390	465	4.75	3.5	1
Color adjectives							
Blue	86	143	459	569	4	3	1
Gray/grey	23	80	471	541	4	3	1
Green	88	116	460	609	5	4	1
Orange	15	23	601	626	6	5	2
Pink	32	48	–	–	4	4	1
Purple	11	13	–	–	6	4	2
Red	125	197	501	585	3	3	1
White	131	365	472	566	5	3	1
Yellow	48	55	537	598	6	4	2
Means	62	116	500	585	4.78	3.67	1.33
Nouns (shapes)							
Circle	19	60	515	591	6	4	2
Diamond	13	8	610	623	7	–	2
Heart	56	173	605	617	5	3	1
Moon	24	60	581	585	4	3	1
Square	32	143	516	610	6	4	1
Star	29	25	574	623	4	3	1
Triangle	28	4	523	597	8	7	3
Means	29	68	561	607	5.71	4	1.57

Notes. SpFr is the spoken frequency taken from the Lancaster University word database. The remaining variables were obtained from the MRC Psycholinguistic Database, University of Western Australia. WrFr is written frequency. Concrt is concreteness. Image is imagability. The last three columns are number of letters, number of phonemes, and number of syllables.

Table 2
Number of trials in the ERP analysis following artifact rejection.

Condition	Min	Max	Mean (SD)	Percentage (%)
Same shapes–color	21	40	30.36 (7.35)	75.9
Same shapes–size	25	40	29.57 (7.74)	73.9
Different shapes–color	24	40	30.00 (7.76)	75.0
Different shapes–size	23	40	31.29 (7.56)	78.2

Note. The artifact rejection included eye blinks, eye movements, and other movements that resulted in a large number of noisy channels. Percentage is per condition.

At the onset of each trial, participants were required to fixate a drift correction dot that appeared in the center of the screen. The trial was initiated by the participant using the spacebar. The display with two objects appeared for 2 s. A fixation cross was then presented in the center of the screen. After a period of 500 ms an auditory instruction referring to one of the objects was played (see Fig. 3 for an example sequence).[2] In order to reduce the effects of eye movements on EEG, we had participants maintain fixation on the cross in the center of the screen during the spoken instruction. This task is similar to an attentional-cueing paradigm (Posner & Cohen, 1984) because it assesses the speed in which participants can orient (covert) attention to an object in the display based on an auditory linguistic description. Therefore, we expected participants to focus attention on the relevant object in the display whilst maintaining fixation on the cross. Many studies have demonstrated that improvements in information processing can occur when attention, but not the eyes, are oriented to a location in visual space (e.g. Cohen, 1981; Ericksen & Hoffman, 1973; Mountcastle, 1978; Posner, 1980; Spence & Driver, 1997). The trial terminated when participants made a button-press response. Participants were seated approximately 60 cm, at head height, from the display monitor, and an experimenter stayed in the testing room during the practice session in order to ensure compliance with the instructions. Participants were reminded before starting the regular session trials that they should maintain fixation on the cross while the auditory instruction was playing, and to try not to blink. Upon completing a block of trials, participants were free to take as long a break as they wanted. The entire experimental session lasted approximately 45 min.

[2] The mean duration of the *look to the* was 558 ms. The mean duration of adjectives for critical trials was 387 ms. The mean duration of nouns for critical trials was 499 ms. Thus, the mean duration of the spoken instructions was 1444 ms.

3. Results

The reaction time data were screened for outliers prior to the analysis. Any data point that was greater than five standard deviations from the mean in each of the four conditions for each participant was considered an outlier. There were four data points that met this criterion and these four values were replaced with the mean for the condition in which they occurred. We utilized a threshold of 300 ms as a cut-off for reaction times; however, there were no trials excluded by this criteria. We also examined correct and incorrect behavioral responses prior to doing the inferential analyses. There were 32 trials that resulted in an incorrect button press, thus participants were correct approximately 99% of the time (see Table 3 for a breakdown of errors by condition). As there were so few trials with an error, we elected to retain them in both the ERP and the reaction time analysis.

3.1. Reaction times

The reaction time results showed that both main effects were significant (display $F(1, 13) = 114.83$, $p < .05$ and modifier $F(1, 13) = 95.04$, $p < .01$). The main effect of display was due to the fact that the displays with two different shapes (or over-described trials) took longer to identify the target compared to displays with the same shapes. The main effect of modifier showed that color adjectives resulted in faster reaction times than did size adjectives. The interaction between display and modifier was also significant $F(1, 13) = 12.24$, $p < .05$ (see Fig. 4). The interaction is driven by the fact that there were larger differences between the two displays with size modifiers compared to the color modifiers. The paired comparison for the size modified trials was significant $t(13) = 8.53$, $p < .05$, and the paired comparison for the color modified trials was also significant $t(13) = 7.14$, $p < .05$. These results are consistent with the hypothesis that over-described referential expressions impair comprehension performance.

3.2. ERP results

The results showed that most of the significant differences occurred between 270 ms and 570 ms past the onset of the adjective, and the scalp distribution suggested that the largest differences were in central, centroparietal, and parietal electrodes (see Fig. 6).[3] The two variables (modifier and display) produced a significant interaction in the early part of this time window (i.e. ~270–420 ms), and there was a main effect of display type in the later part of the window (i.e. 450–570 ms). The results of the inferential tests

[3] Figs. 6 and 7 have positive plotted upwards and negative downwards. Panel B (size results) suggests a late positivity emerging approximately 800 ms post-adjective onset. However, there were no significant differences by the criteria described in the Methods. An analysis of individual participant means suggested that a minority of participants were responsible for this late P600-like effect.

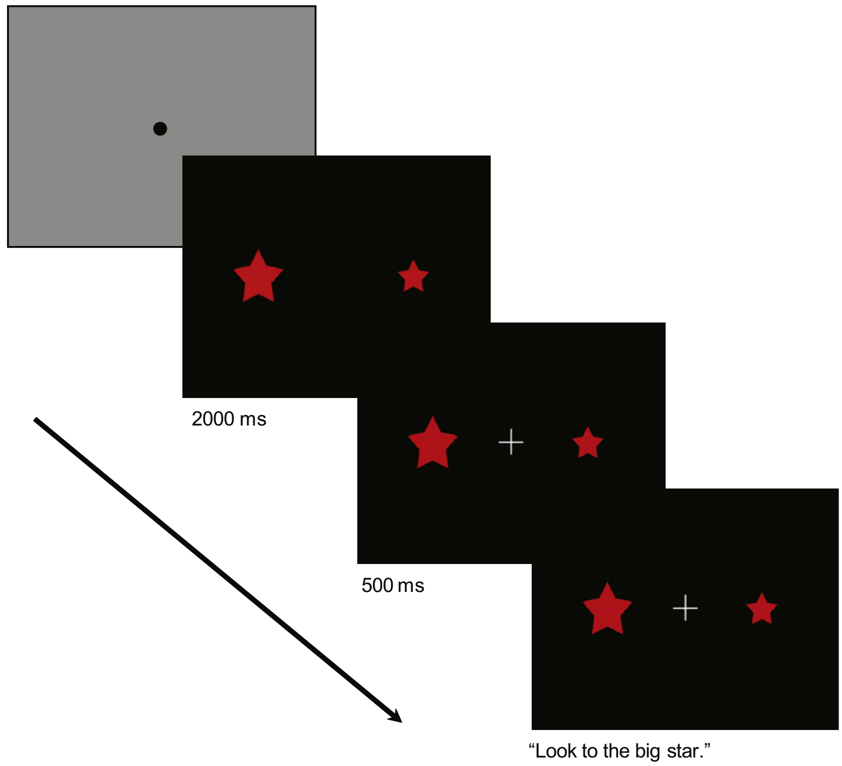

Fig. 3. Example trial sequence and timing of events within a trial.

Table 3
Number of trials and reaction time results for error trials.

Condition	# With error	Mean RT (SD)
Same shapes–color	14 (2.50%)	754 (253)
Same shapes–size	4 (0.71%)	1017 (169)
Different shapes–color	10 (1.79%)	947 (292)
Different shapes–size	4 (0.71%)	1242 (622)

Note. Percentages are per condition.

conducted on each region of interest in which there were significant differences are presented in Tables 4 and 5. As can be seen from these tables, the results are remarkably consistent across regions of interest.

As a final analysis, we collapsed across the significant regions of interest (i.e. only those reported in Tables 4 and 5). The grand averaged results are presented in Fig. 7. Panel A of Fig. 7 shows the results from the earlier time window (270–420 ms). A 2 × 2 repeated measures ANOVA revealed that the two variables interacted $F(1, 13) = 8.22$, $p < .05$. We followed up the significant interaction with paired samples *t*-tests, focusing particularly on the effect of display. Comparing the trials with size modifiers, the difference was not significant ($p > .20$). However, the two conditions with color modifiers showed a marginal difference $t(13) = 1.88$, $p = .08$. Thus, the significant interaction is driven by the fact that the color modifiers were more positive compared to size modifiers, and the color condition with two different objects (over-described) was more negative than the color condition with the same objects. The grand averaged results in the later time window (450–570 ms) are presented in Fig. 7, panel B. Here the results showed a main effect of display $F(1, 13) = 9.35$, $p < .05$, and the displays with different objects (over-described instructions) were more negative than the displays that had two objects of the same type. Thus, over-descriptions result in a negativity in a time window suggestive of a N400 effect.

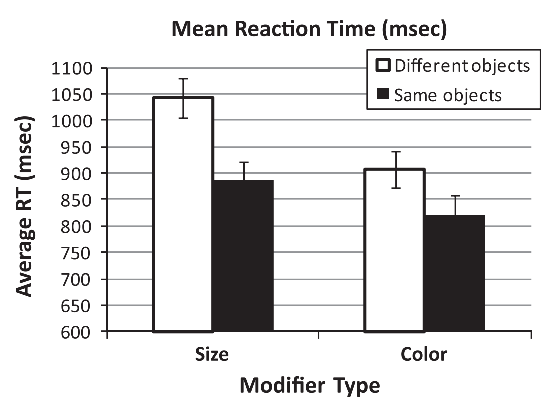

Fig. 4. Reaction time results from the onset of the adjective for each of the four conditions in the experiment. Error bars show the standard error of the mean.

3.3. Filler trials

In our critical trials with two shapes of the same type, we only manipulated a single feature in the display, which makes the form of the referring expression somewhat predictable. For example, with panel A of Fig. 1, participants can predict that they will hear a size adjective (i.e. *big*, *large*, *small*, or *little*), and with panel C, they can predict either *red* or *blue*. Recall that we did include 80 filler trials that had both size and color manipulated in the same display (see Fig. 2, panel A). Half of these trials had a size adjective and half a color adjective. Crucially, with these displays it is impossible to predict which type of modifier will be in the instruction. Therefore, we analyzed the reaction time data from these trials in order to compare the results to trials in the main analysis that had two objects of the same type (i.e. displays like Fig. 1, panels A and C). Any difference between those in the main analysis and the 80 filler trials, must be due to differences in predictability.

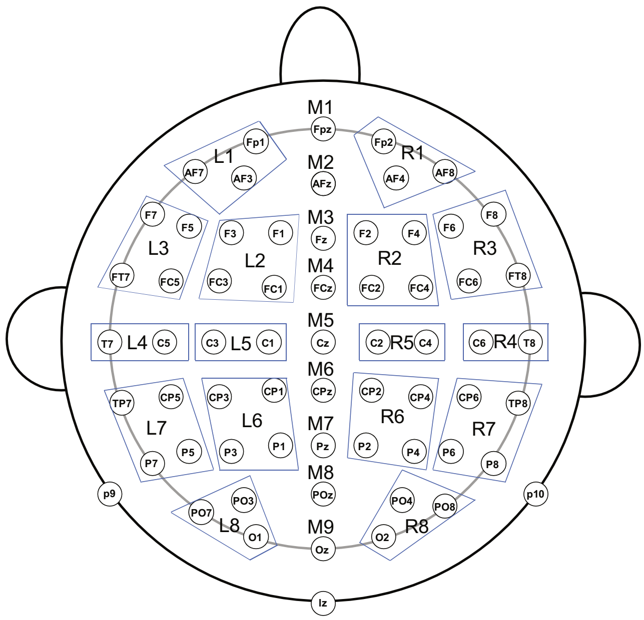

Fig. 5. Scalp configuration of the 25 regions of interest. There are eight on the left, nine on the midline, and eight on the right.

The reaction time means for these 80 filler trials fell in between the ones from the main analysis: size (M = 942 ms, SEM = 36.9) and color (M = 864 ms, SEM = 35.0). These trials are similar to all of the trials in the main analysis, in that, the target object is uniquely identifiable at the modifier. The difference for the size modified trials with two objects of the same type (filler minus critical) was 56 ms, which was significant $t(13) = 3.23$, $p < .01$. The difference between the color modified trials with two objects of the same type (filler minus critical) was 43 ms, which was also significant $t(13) = 2.44$, $p < .05$. Thus, the predictability benefit was approximately 50 ms. We return to the issue of predictability in the General Discussion.

3.4. Summary

There were substantial slowdowns in reaction time when a referring expression contained an unnecessary modifier. In addition, both size and color modifiers elicited a negative ERP component that peaked at approximately 500 ms past the onset of the adjective. The timing and distribution of this effect is similar to an N400, suggesting that extra information (i.e. an additional adjective modifier) combined with a visual display that contains two different objects results in an integration problem. Our results showed that the negativity began to emerge sooner with color modifiers than it did with size modifiers. These time course differences in the ERP signal map onto the reaction time results as there was a main effect of modifier on reaction times in which the color modifiers were processed ~100 ms faster than the size modifiers (for a similar result, see Engelhardt, Xiang, & Ferreira, 2008).

4. Discussion

The primary goal of this study was to determine whether over-described referring expressions facilitate or impair comprehension performance. The results from both behavioral and electrophysiological measures indicated that there are very clear processing impairments when extra (or unnecessary) information is included. Our findings, therefore, support the predictions made by Grice (1989) concerning the second part of the Maxim of Quantity. Namely, that listeners are confused by the inclusion of extra information.

The behavioral results showed that reaction times were much slower with over-described instructions compared to instructions that had the "appropriate" amount of information. The difference was approximately twice as large for size (156 ms) as it was for color (86 ms). These reaction time differences are substantial, because in all four conditions of the experiment, the target is uniquely identifiable at the modifier. The slowdown that we observed suggests that participants generate a quantity-based expectation or prediction. This is not surprising, as several other studies in the literature have shown that comprehension processes to some extent engage active prediction concerning what type of information is likely to follow (Altmann & Kamide, 1999; Lau, Stroud, Plesch, & Phillips, 2006; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Staub & Clifton, 2006). When participants are presented with two different objects, it seems that they have an expectation for an unmodified noun phrase, and when this expectation is violated with an over-description, comprehension is slowed. On the other hand, when there are two objects of the same type, participants know that a modifier is required, and thus they expect a size or color modifier depending on the display.

Related to the issue of predictability, we analyzed a subset of the filler trials that contained two objects of the same type, and had both size and color manipulated (see Fig. 2, panel A). These trials are interesting because they give insight into the processing benefits of being able to predict a particular class of modifier. Essentially, with these filler trials participants cannot predict whether the instruction will contain a size or a color modifier, as it could be either. When we compared the reaction times from these trials to the ones in the main analysis that had two objects of the same type (Fig. 1, panels A and C), we found that participants were faster with displays that had only one feature manipulated at a time. That is, participants are faster when they can predict which type of modifier they will hear. However, these filler trials also show that there is some additional cost when the display contains different objects and a pre-nominal modifier is heard (i.e. when the instruction is over-description). This additional cost was 100 ms in the size modified trials and 43 ms in the color modified trials, which shows that the difference between displays with the same shapes vs. different shapes is not entirely due to prediction.

The ERP results revealed that instructions with an unnecessary modifier triggered a centroparietal negativity that emerged earlier with color modifiers compared to size modifiers. The timing of this negativity, and the scalp distribution are very similar to those commonly reported in N400 studies, except that the effect occurred over a larger region of left hemisphere (i.e. L4, L5, L6, and L7). Many of the previous N400 studies, with visual presentation, showed larger effects over the right hemisphere, and consistent with this, we observed statistically greater differences with the paired comparisons in the right hemisphere compared to the left hemisphere (i.e. R6 and R7). However, these differences were confined to a smaller region of the scalp.

In the Introduction, we reviewed three of the most prominent ERP components that have been associated with language comprehension (i.e. N*ref*, N400, and P600). Our results showed significant negativity in a time window much like an N400. There are two main views on the N400. The first is that it represents a semantic-integration problem between a word and the previous context. The second is that it reflects the degree of difficulty accessing a word from long term memory. When words are highly predictable, the comprehension system can pre-activate particular lexical items in long term memory, which makes lexical access operations less effortful. We believe that our results are partially compatible with both of these views. It could be the case that the negative ERP potentials and the slower reaction times with the over-described

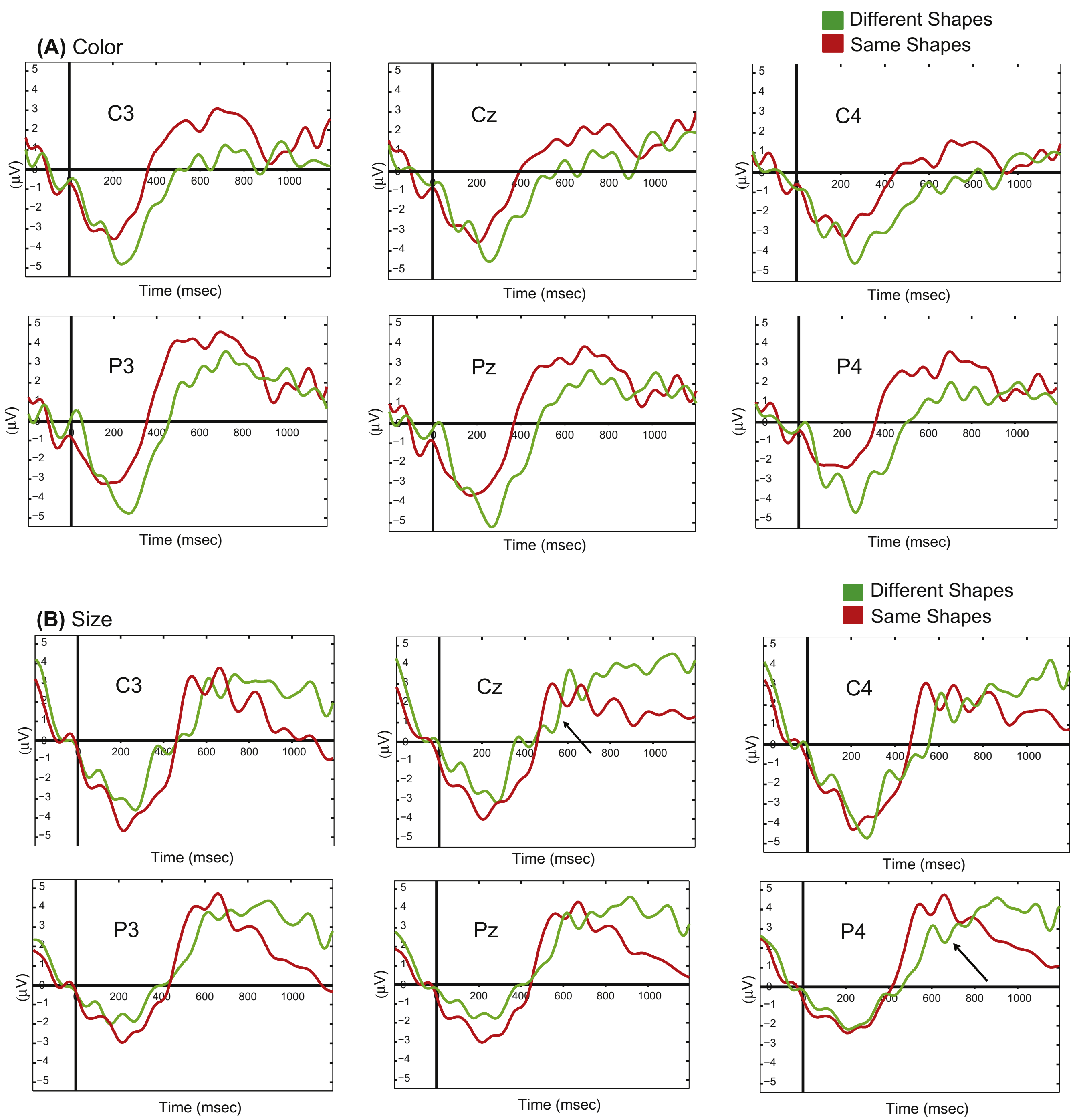

Fig. 6. Example electrodes from central and parietal scalp regions. Panel A shows color results and panel B shows size results.

instructions, that we observed, is due to more difficult integration of the spoken modifier with a visual context that does not support its use. On the other hand, a substantial proportion of the difference in reaction times must be due to facilitation in the conditions where a modifier is required. Recall that the filler trials in which both features were manipulated resulted in reaction times that fell in between those from the main analysis. Therefore, it seems that part of our effect is due to predictability when the display contains two objects of the same type and only one feature is manipulated. However, at the same time, there is also some additional impairment from the extra modifier when the display contains two different objects. Our results therefore, indicate that the processing difficulty associated with over-descriptions is likely a semantic-integration or lexical-access problem. There was no evidence of an *Nref* component, which is observed with referential under-description (Van Berkum et al., 1999). We also did not observe significant differences in the P600 time window, which indirectly indicates that over-descriptions are not registered as a word-category or more general syntactic problem (Friederici et al., 1996; Osterhout & Holcomb, 1992).

The overall goal of this study was to adjudicate between the two competing views of the comprehension of over-descriptions.

Table 4
Analyses of variance for early interaction between the two variables: modifier type × display type.

Region of interest	Interaction	Display type	Modifier type	T1	T2
Left					
L4 (180–480)	$F = 11.26$, $p = .01$	$F = .29$, $p = .60$	$F = 3.91$, $p = .07$	$t = 1.82$, $p = .09$	$t = -1.72$, $p = .11$
L5 (270–450)	$F = 6.32$, $p = .03$	$F = .18$, $p = .68$	$F = 3.40$, $p = .09$	$t = 1.63$, $p = .13$	$t = -1.56$, $p = .14$
L6 (270–450)	$F = 9.74$, $p = .01$	$F = .46$, $p = .51$	$F = 8.37$, $p = .01$	$t = 1.87$, $p = .09$	$t = -1.30$, $p = .22$
L7 (210–480)	$F = 11.19$, $p = .01$	$F = .36$, $p = .56$	$F = 8.57$, $p = .01$	$t = 1.75$, $p = .10$	$t = -1.16$, $p = .27$
Midline					
M5 (270–450)	$F = 4.51$, $p = .05$	$F = .31$, $p = .59$	$F = .67$, $p = .43$	$t = 1.63$, $p = .13$	$t = -1.31$, $p = .21$
M6 (270–420)	$F = 5.83$, $p = .03$	$F = .37$, $p = .55$	$F = 2.02$, $p = .18$	$t = 1.77$, $p = .10$	$t = -1.49$, $p = .16$
M7 (180–450)	$F = 7.54$, $p = .02$	$F = .45$, $p = .52$	$F = 3.43$, $p = .09$	$t = 1.85$, $p = .09$	$t = -1.50$, $p = .16$
M8 (180–420)	$F = 11.84$, $p = .00$	$F = .34$, $p = .57$	$F = 2.93$, $p = .11$	$t = 1.80$, $p = .10$	$t = -1.73$, $p = .11$
Right					
R6 (270–450)	$F = 6.78$, $p = .02$	$F = 1.26$, $p = .28$	$F = 2.11$, $p = .17$	$t = 1.98$, $p = .06$	$t = -.60$, $p = .56$
R7 (360–450)	$F = 7.11$, $p = .02$	$F = 3.67$, $p = .08$	$F = 1.89$, $p = .19$	$t = 2.45$, $p = .03$	$t = .28$, $p = .79$

Note. T1 is the paired comparison for the conditions with color modifiers. T2 is the paired comparison for the conditions with size modifiers.

Table 5
Analyses of variance for the late main effect of display type.

Region of interest	Interaction	Display type	Modifier type	T1	T2
Left					
L4 (480–570)	$F = .37$, $p = .56$	$F = 9.33$, $p = .01$	$F = .86$, $p = .37$	$t = 2.16$, $p = .05$	$t = 2.16$, $p = .05$
L5 (480–570)	$F = .04$, $p = .84$	$F = 6.60$, $p = .02$	$F = .01$, $p = .93$	$t = 1.60$, $p = .14$	$t = 1.69$, $p = .11$
L6 (450–570)	$F = .17$, $p = .69$	$F = 6.88$, $p = .02$	$F = .53$, $p = .48$	$t = 1.79$, $p = .09$	$t = 1.71$, $p = .11$
L8 (450–600)	$F = .25$, $p = .62$	$F = 10.09$, $p = .01$	$F = .01$, $p = .92$	$t = 1.47$, $p = .17$	$t = 2.66$, $p = .02$
Midline					
M5 (450–570)	$F = .16$, $p = .70$	$F = 5.87$, $p = .03$	$F = .01$, $p = .93$	$t = 1.66$, $p = .12$	$t = 1.17$, $p = .26$
M6 (450–570)	$F = .10$, $p = .75$	$F = 7.38$, $p = .02$	$F = .02$, $p = .91$	$t = 1.87$, $p = .09$	$t = 1.57$, $p = .14$
M7 (450–570)	$F = .18$, $p = .68$	$F = 8.32$, $p = .01$	$F = .01$, $p = .98$	$t = 1.89$, $p = .08$	$t = 1.62$, $p = .13$
M8 (450–600)	$F = .01$, $p = .97$	$F = 9.44$, $p = .01$	$F = .06$, $p = .81$	$t = 1.58$, $p = .14$	$t = 1.67$, $p = .12$
Right					
R5 (450–570)	$F = .07$, $p = .80$	$F = 6.14$, $p = .03$	$F = .39$, $p = .54$	$t = 1.71$, $p = .11$	$t = 1.78$, $p = .10$
R6 (450–570)	$F = .02$, $p = .89$	$F = 9.30$, $p = .01$	$F = .20$, $p = .66$	$t = 1.86$, $p = .09$	$t = 2.19$, $p = .05$
R7 (450–570)	$F = .00$, $p = .99$	$F = 8.10$, $p = .01$	$F = .73$, $p = .41$	$t = 1.78$, $p = .10$	$t = 2.96$, $p = .01$
R8 (390–600)	$F = .53$, $p = .48$	$F = 6.47$, $p = .02$	$F = .01$, $p = .92$	$t = 1.96$, $p = .07$	$t = 2.56$, $p = .02$

Note. T1 is the paired comparison for the conditions with color modifiers. T2 is the paired comparison for the conditions with size modifiers.

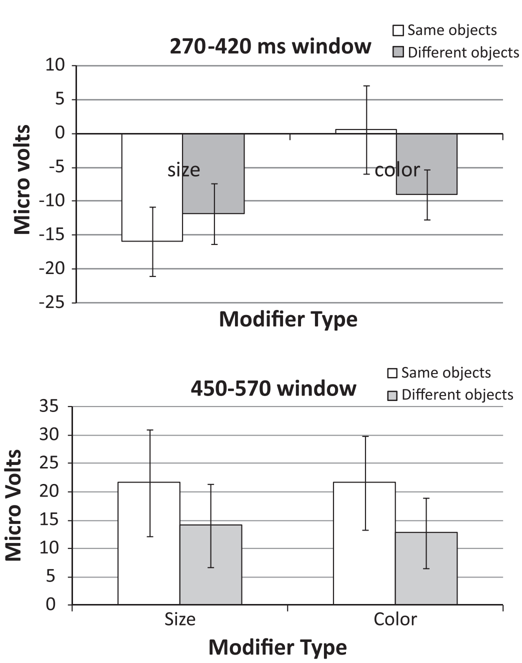

Fig. 7. Grand averaged ERP results. Panel A shows the results for early time window, and panel B shows the results for the late time window. Note that positive is plotted up and negative is down.

There has been a lot of variability in the task requirements and types of modifiers that have been used in previous studies to address the comprehension of over-descriptions. Some have argued that over-descriptions are problematic to comprehension and others have argued that over-descriptions are beneficial. The latter of these positions has been predominantly argued for by Arts and colleagues. In a recent paper, they found that after reading an object description with extra information that participants were faster to identify an object on a subsequent screen. The benefits they observed ranged from 60 to 100 ms depending on the type of extra information that was included (e.g. size, shape, color, or spatial location). In contrast, in Engelhardt et al. (2006) and in the studies by Sedivy and colleagues, participants had their eye movements monitored while they moved real objects in a workspace or viewed objects on a computer screen. In these studies, participants heard instructions referring to objects that were co-present (i.e. always in view).

Therefore, there are two main differences between these studies. The first is the modality of presentation: written vs. auditory. The second is whether the objects are co-present with the linguistic description. In the Arts et al. paradigm, participants build a mental representation of an object, which they must then identify in a subsequently presented array. In contrast, in the current work, the objects are presented prior to the linguistic description and remain in view until the participant has made a response. Therefore, participants can incrementally parse the linguistic description, and immediately combine the linguistic information with the available visual context. However, the grand mean for reaction times across all conditions in our experiment was ~900 ms and the grand mean in Arts et al. (2011) was ~1100 ms, which suggests that there are not major processing differences based on whether the objects are co-present or appear after the linguistic description. We do not want to make the claim that either situation is an inherently better test of the hypothesis that we set out to test, but instead, we hope that the current study will lead increased interest and future studies to tease apart the various factors that influence how listeners deal with extra information (i.e. over-descriptions).

There are two minor issues that we feel should be considered when interpreting the findings from this study as a whole. The first is that we deliberately tested simple visual contexts, which affords some degree of predictability. It remains an open issue whether we would observe reaction time slowdowns and N400-like effects with complex displays, such as real-world scenes. It is likely that the ability to predict the presence/absence of a modifier with more complex displays would be limited by general cognitive processing constraints, such as limitations on visual attention and/or working memory.

The second caveat also concerns the role of predictability. Recall that in the color condition with two objects of the same shape, participants can essentially pre-activate or predict one of two lexical items. In Fig. 1, panel C, the modifier must be either *red* or *blue*. In contrast, with the size displays (Fig. 1, panel A), there is less predictability because four lexical items are possible *little, small, large*, or *big* (see Table 1). This difference might go some way towards explaining the general processing advantage of color compared to size. However, there are several other reasons why color and size might be processed differently. The first is that size is a relative modifier, which means that the value of "big", for example, is only defined with respect to some comparison class (Ferris, 1993; Kamp, 1975; Siegel, 1980). Whereas, color is an inherent property of an object, which means that the value is fixed. A second possibility is lexical variables, which we have presented in Table 1. Size modifiers had higher frequencies than color modifiers ($p < .05$); however, higher frequency words are typically processed more easily than lower frequency words (Rayner, 1998). Concreteness and imagability, on the other hand, were significantly higher ($p < .05$) for color compared to size, which could contribute to the general processing differences that we observed. Whatever the reason, the processing differences between size and color modifiers does not influence our main conclusions because the effect of over-description was the same for both classes of adjectives.

5. Conclusion

In summary, the ERP results showed an N400 effect when participants heard a referential expression that contained a modifier and the visual display had two different objects (i.e. an over-description). The N400 effect emerged earlier with color modifiers compared to size modifiers, and in the 420–570 ms time window, we only observed a main effect of modifier type in which the over-described instructions were more negative. The reaction time data showed that over-described instructions resulted in substantially longer reactions times compared to instructions that had the appropriate amount of information. Thus, our results support the assumption that too much information is detrimental to comprehension performance, which is consistent with second part of the Maxim of Quantity.

Acknowledgments

The authors would like to thank Laura Speed, Yelda Semizer, and Sandra M. Andraszewicz for the running experiment and the preparing stimuli. This work was supported by ESRC Grant RES-062-23-0475 awarded to Fernanda Ferreira.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.
Arts, A. (2004). Overspecification in instructive texts. Doctoral dissertation, Tillburg University.
Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics, 43*, 361–374.
Belke, E. (2006). Visual determinants of preferred adjective order. *Visual Cognition, 14*, 261–294.
Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). *Processing syntax and morphology: A neurocognitive perspective*. Oxford University Press.
Cohen, Y. (1981). Internal and external control of orienting. Doctoral dissertation, University of Oregon.
Davies, C., & Katsos, N. (2009). *Are interlocuters as sensitive to over-informativeness as they are to under-informativeness? PRE-CogSci workshop: Production of referring expressions*. Austin, TX: Cognitive Science Society.
Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent components analysis. *Journal of Neuroscience Methods, 134*, 9–21.
Demiral, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2008). On the universality of language comprehension strategies: Evidence from Turkish. *Cognition, 106*, 484–500.
Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition, 11*, 159–184.
Engelhardt, P. E. (2008). Generalized implicatures: Do uniformative environments eliminate default interpretations? Doctoral dissertation, Michigan State University.
Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean maxim of quantity. *Journal of Memory and Language, 54*, 554–573.
Engelhardt, P. E., Xiang, M., & Ferreira, F. (2008). Anticipatory eye movements mediated by word order constraints. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 951–957). Austin, TX: Cognitive Science Society.
Ericksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception and Psychophysics, 14*, 155–160.
Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*, 469–495.
Ferris, C. (1993). *The meaning of syntax: A study of adjectives in English*. London & New York: Longman.
Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*, 78–84.
Friederici, A. E., Hahne, A., & Meklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1219–1248.
Grice, P. (1975). Logic and Conversation. In P. Cole, J. Morgan (Eds.), *Syntax and semantics: Speech acts*. (Vol. III, pp. 41–58).
Grice, P. (1989). *Studies in the ways of words*. Cambridge, MA: Harvard University Press.
Grodner, D., & Sedivy, J. (2011). The effects of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (Eds.), *The processing and acquisition of reference*. Cambridge MA: MIT Press.
Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 273–316). Oxford: Oxford University Press.
Kamp, H. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal semantics for natural languages* (pp. 123–155). Cambridge, UK: Cambridge University Press.
Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 4*, 463–470.
Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.
Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 61–91). Oxford, UK: Elsevier Inc.
Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews: Neuroscience, 9*, 920–933.
Lau, E. F., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language, 98*, 74–88.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT.
Maes, A., Arts, A., & Noordman, L. (2004). Reference management in instructive discourse. *Discourse Processes, 37*, 117–144.
Mountcastle, V. B. (1978). Brain mechanisms of directed attention. *Journal of the Royal Society of Medicine, 71*, 14–28.

Nieuwland, M. S., & Van Berkum, J. A. (2008). The neurocognition of referential ambiguity in language comprehension. *Language and Linguistics Compass, 2*, 603–630.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics, 27*, 89–110.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 2–25.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bowhuis (Eds.), *Attention and Performance X* (pp. 531–556). Hove, UK: Lawrence Erlbaum Associates Ltd.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Rugg, M. D., & Coles, M. G. H. (1997). *Electrophysiology of mind* (Vol. 25). Oxford, UK: Oxford University Press.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research, 32*, 3–23.

Sedivy, J. C. (2006). Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in online language interpretation. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to world-situated language use: Bridging the language-as-product and language-as-action traditions (learning, development, and conceptual change)* (pp. 345–364). Cambridge MA: MIT Press.

Sedivy, J. (2007). Implicature during real time conversation: A view from language processing research. *Philosophy Compass, 2*, 475–496.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual interpretation. *Cognition, 71*, 109–147.

Siegel, M. (1980). *Capturing the adjective*. New York: Garland.

Sonnenschein, S. (1984). The effect of redundant communication on listeners: Why different types have different effects. *Journal of Psycholinguistic Research, 13*, 147–166.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception and Psychophysics, 59*, 1–22.

Staub, A., & Clifton, C. Jr., (2006). Syntactic prediction in language comprehension: Evidence from either…or. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 425–436.

Van Berkum, J. A. (2008). Understanding sentences in context. *Current Directions in Psychological Science, 17*, 376–380.

Van Berkum, J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*, 235–248.

Van Berkum, J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*, 657–671.

Van Berkum, J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*, 158–171.