**Article:**

Davies, C orcid.org/0000-0001-9347-7905 and Kreysa, H (2017) Looking at a contrast object before speaking boosts referential informativeness, but is not essential. Acta Psychologica, 178. pp. 87-99. ISSN 0001-6918

https://doi.org/10.1016/j.actpsy.2017.06.001

*This is an author-accepted version of the manuscript accepted by Acta Psychologica on 06 June 2017.*

# Looking at a contrast object before speaking boosts referential informativeness, but is not essential.


## Abstract

Variation in referential form has traditionally been accounted for by theoretical frameworks focusing on linguistic and discourse features. Despite the explosion of interest in eye tracking methods in psycholinguistics, the role of visual scanning behaviour in informative reference production is yet to be comprehensively investigated. Here we examine the relationship between speakers' fixations to relevant referents and the form of the referring expressions they produce. Overall, speakers were fully informative across simple and (to a lesser extent) more complex displays, providing appropriately modified referring expressions to enable their addressee to locate the target object. Analysis of contrast fixations revealed that looking at a contrast object boosts but is not essential for full informativeness. Contrast fixations which take place immediately before speaking provide the greatest boost. Informative referring expressions were also associated with later speech onsets than underinformative ones. Based on the finding that fixations during speech planning facilitate but do not fully predict informative referring, direct visual scanning is ruled out as a prerequisite for informativeness. Instead, pragmatic expectations of informativeness may play a more important role. Results are consistent with a goal-based link between eye movements and language processing, here applied for the first time to production processes.


**Keywords**: Reference; Speech production; Informativeness; Pragmatics; Eye movements.

## 1. Introduction

A large body of research on reference has documented context-dependent variation in speakers' referring expressions, ranging from minimal choices such as null and pronominal forms up to more explicit modified noun phrases (for a review see Davies and Arnold, in press). For example, when introducing a referent into the discourse, a speaker is likely to say 'an apple', whereas if the apple had recently been mentioned, a pronoun would be more likely.

Variation in referential choice has traditionally been accounted for by theoretical frameworks focusing on how referential expressions are constrained by the linguistic discourse context (Ariel, 1990, 2001; Chafe, 1976, 1994; Gundel, Hedberg and Zacharski, 1993; Gordon and Hendrick, 1998; Grosz, Joshi, and Weinstein, 1995), which affects the referents' information status (see Arnold, Kaiser, Kahn and Kim, 2013, for a review). The current paper focuses on modified noun phrases; specifically on speakers' choice of whether or not to include a prenominal adjective. This type of modifier inclusion has been found to vary across studies, mediated by a number of factors such as display density (Arnold and Griffin, 2007; Koolen, Krahmer and Swerts, 2015); number and type of attributes held by the target referent (Koolen, Gatt, Goudbeek and Krahmer, 2011; Mangold and Pobel, 1988; Tarenskeen, Broesema and Geurts, 2015; van Gompel, Gatt, Krahmer and van Deemter, 2014); discourse goals (Arts, Maes, Noordman and Jansen, 2011a; Maes, Arts and Noordman, 2004); use of ground information (Brennan and Clark, 1996; Nadig and Sedivy, 2002), and individual differences (Davies and Katsos, 2010; Hendriks, 2016). Here we focus on speakers' visual scanning behaviour as a predictor of informativeness.

In the paradigm used in this paper, inclusion or exclusion of a modifier in a referring expression affects that expression's informativeness. Referential informativeness depends on the relationship between the context (including visual and discourse features) and the referring expression. Thus, we define informativeness as a property of expressions within their contexts, such that more informative expressions are those that match a smaller set of candidate referents. In line with previous work in this area (Engelhardt, Bailey and Ferreira, 2006; Davies and Katsos, 2010, i.a.), we adopt a three-way taxonomy in which referring expressions can be optimally informative (e.g. 'the small apple' to refer to one of a pair of apples contrasting in size), underinformative ('the apple' in the same context), or overinformative ('the small green apple' when there is only one apple). The current study focuses on speakers' visual interrogation of a scene before producing informative vs. underinformative expressions.

From a processing perspective, the production of an expression to refer to an object in a referential communication paradigm is a multi-stage process requiring a variety of skills. Firstly, the speaker must attend to the target referent to perceive its characteristics, and visually scan the surrounding scene to check for same-category competitors. This allows the speaker to identify such competitors and consider which distinguishing features should be encoded into the utterance in order to avoid producing an ambiguous referring expression (cf. Grice's Quantity maxim; 1975/1989). The speaker must also assess the accessibility of the target referent for themselves and for their addressee: is it focused or shared in the current linguistic or extralinguistic context? The cooperative speaker should then integrate this information into a coherent and fully informative referring expression. Classic theories of accessibility and reference production have accounted for the latter stages of this process (see Arnold, 2008, for a review), and while these have been extensively empirically investigated (Keysar, Barr, Balin and Brauner, 2000; Brown-Schmidt, Gunlogson and Tanenhaus, 2008; Hanna, Tanenhaus and Trueswell, 2003; Heller, Grodner and Tanenhaus, 2008, i.a.), the earlier, speaker-oriented stages in reference production have received less attention (though see Bock, Irwin, and Davidson, 2004; Bock, Irwin, Davidson, and Levelt, 2003; Griffin and Bock, 2000; Kuchinsky, Bock, and Irwin, 2011, for discussions of early processes of speech production in general). This paper aims to redress the balance by focusing on the processes at work during speech planning in referential communication.

The existing literature on variation in referential informativeness has focused on both bottom-up influences such as features of the referent, and top-down factors such as the use of common ground. However, it has not yet comprehensively addressed the question of how *visual scanning behaviour* might affect referential choice (appealed for by Deutsch and Pechmann, 1982: 177, and documented as part of a wider study by Brown-Schmidt and Tanenhaus, 2006). Intuitively, if speakers do not complete a full scan of the visual environment, they may not realise that there are objects co-present that belong to the same category as the target and must be distinguished from it. Thus, they risk being underinformative. Pechmann (1989: 98) suggested that incomplete visual scanning might be a reason for failures in informativeness, when '[…] the speaker initially pays attention to the target object without seriously considering the context'. He explained that due to the incremental nature of speech processing, such behaviour may lead to overspecification if speakers start articulating their utterance before they have scanned the whole display and deduced the distinctive feature(s) of the target referent. Such behaviour could render the early part of the referring expression noncontrastive, e.g. 'the white circle', where there is only one circle in the display. Although Pechmann's work focused on rates of over- rather than under-specification, the same process could also plausibly result in ambiguous referring expressions, where a speaker mentions features of the target referent which fail to distinguish it from its competitors, e.g. 'the white circle' in a context containing a large and a small white circle. Further (though indirect) evidence for a close relationship between visual scanning and infelicitous informativeness comes from studies finding higher rates of overinformativeness when there are more competitors to scan for discriminating features (Koolen, Krahmer and Swerts, 2015; Mangold and Pobel, 1988). Due to the availability of eye tracking technology, it is now possible to directly examine the relationship between visual scene interrogation and the production of referential attributes. We investigate this link by examining eye movements to target and contrast objects before the articulation of fully informative and underinformative referring expressions.

Previous research on pragmatic informativeness has concentrated on comprehension in investigating the interaction of reference and eye movements. Classic work using the visual world paradigm has shown that referential context is pivotal in the interpretation of temporary referential ambiguities (Chambers, Tanenhaus and Magnuson, 2004; Sedivy, Tanenhaus, Chambers and Carlson, 1999; Sedivy, 2003; Tanenhaus, Spivey-Knowlton, Eberhard and Sedivy, 1995; Trueswell, Sekerina, Hill and Logrip, 1999). This focus on reference comprehension is influenced by the wider psycholinguistic tradition of measuring eye movements in language comprehension (see Huettig, Rommers and Meyer, 2011, and Altmann, 2011 for reviews). Although there has been comparatively little research into eye movements during language production, studies published throughout the 90s and early 2000s furthered our understanding of the relationship of eye movements to speech planning and articulation, e.g. the time-locking of eye movements and speech (Griffin and Bock, 2000) and the influence of word frequency and visual clarity on pre-articulatory viewing times (Meyer, Sleiderink and Levelt, 1998; see Meyer, 2004, and Griffin, 2004 for reviews). This research provides important foundations for the current study, i.e. that fixations to objects typically precede reference to them, and more broadly, that eye movements convey information about speech planning processes that precede the onset of an utterance as well as about those which occur during articulation. The current study extends existing work by using eye tracking to study the production of pragmatic informativeness in tightly controlled referential forms.

Our study extends three recent papers on eye movements and informativeness by analysing fixations according to the informativeness of referring expressions. Firstly, Rabagliati and Robertson (2016, exp. 1 and 1a) monitored speakers' fixations to target and contrast objects in scenes containing lexical ambiguity (e.g. a baseball bat and an animal bat) and non-linguistic ambiguities (e.g. a red car and a yellow car), finding that adult speakers proactively monitored for non-linguistic ambiguity before articulating their referring expressions, as well as in a post-naming monitoring phase. Secondly, Vanlangendonck, Willems, Menenti and Hagoort (2016) monitored speakers' fixations to target and contrast objects in common vs. privileged ground in order to test between competing accounts of common ground use. Speakers were found to initially fixate the target object, then consider other objects in the array. Ultimately, the number of fixations to the contrast object during the analysed temporal region (i.e. from the highlighting of the target object to utterance onset) was low, though since the authors did not analyse fixation patterns during the preview region it is unclear whether speakers were relying on previous fixations while planning their utterance (as acknowledged by Vanlangendonck et al., 2016: 749). We extend this work by presenting this precise analysis. Thirdly, Brown-Schmidt and Tanenhaus (2006) explored the relationship between the timing of a first fixation to a contrast object and the form of subsequent referring expressions in a referential communication game. The production of a fully informative referring expression was more likely if the speaker had fixated a contrast referent. However, in their exp. 1, which used relatively simple displays comprising geometric shapes and simple images, 68% of utterances were informative even *without* a contrast fixation. In their exp. 2, which used more naturalistic scenes and additional referents within each display, 19% of utterances were informative without a contrast fixation. Brown-Schmidt and Tanenhaus's (2006) findings thus provide evidence against a mechanistic account of reference in which contrast objects must be checked and assessed, either before or after the onset of the utterance (Meyer, Sleiderink and Levelt, 1998; Griffin and Bock, 2000). Instead, it seems that speakers can indeed be fully informative without fixating contrast objects. We extend this work by analysing relative fixations to contrast and target objects, and by manipulating set size within a single experiment.

These production findings from Rabagliati and Robertson (2016), Vanlangendonck et al. (2016), and Brown-Schmidt and Tanenhaus (2006) all accord with a *goal-based* linking hypothesis that describes the relationship between eye movements and language processing in language comprehension (Salverda, Brown and Tanenhaus, 2011). On this account, different types of representations are involved in mapping speech to a scene, depending on the viewer's current task. Eye movements are assumed to reflect task-specific visual processes (as well as general-purpose ones), in which locations most relevant to the task at hand are more likely to be fixated. Evidence comes from Altmann and Kamide's (1999) demonstration of anticipatory effects in incremental language processing. Although this classic study has been cited intensively, few have commented on the clear effects of experimental task, with earlier and more looks to the target during the verb region in a sentence verification task than in a more passive look-and-listen task. Further support for the goal-based view comes from Brown-Schmidt and Tanenhaus (2008), who recorded fewer looks to task-irrelevant objects than to relevant ones in a referential communication game, even when the former matched the referring expression heard by the addressee. That is, on hearing 'Put the green block above the red block', participants were more likely to look at a red block with an empty space above it (rendering it compatible with the task) than at a red block without a vacant space above. Both blocks were linguistic matches for the referring expression, but looking behaviour was clearly mediated by the extralinguistic referential context. The current study aims to test one of the predictions of the goal-based account, i.e. that the referent that is most relevant for the task at hand will be the one that receives the most visual attention. Notably, it does so using a language production paradigm in an interactive setting.

### 1.1 The current study: Task and hypotheses

We measured speakers' referential informativeness and their accompanying eye movements as they completed an interactive referential task. Participants saw arrays of four or eight objects, containing a singleton target object (e.g. a ball) or a target object accompanied by a contrast mate (e.g. a large and a small ball). They then told their addressee (who could see the same array without the target highlighted) to click on the target (see Figs. 1 and 2 for example displays).

Despite intuitions that looking at an object is a prerequisite for referring to it informatively, there is some evidence to suggest that even target objects do not necessarily require a direct fixation in order to be referred to correctly (Dobel, Gumnior, Bölte and Zwitserlood, 2007; Griffin 2004: 231). This evidence has yet to be reliably extended to *contrast objects*. Contrast objects are less salient than the target object in referential communication tasks, but are still highly relevant for the goal of felicitous referring. The goal-based linking hypothesis predicts that the experimental task will mediate fixation patterns, i.e. speakers will be attracted to the most relevant referent for the task at hand. To test this hypothesis, we measure relative attention on contrast and target. Further, previous research suggests that objects can be processed extrafoveally and/or in parallel when a referring expression is very easy to generate (e.g. a pronoun) or when an object is highly recognisable (Meyer et al., 1998; Morgan and Meyer, 2005), but it is not yet clear whether this holds for discourse contexts where speakers can only be fully informative if they use information from a contrast object. Regarding the relationship between scanning behaviour and speech onset time, and assuming a serial view of speech planning (Levelt, 1989), more comprehensive pre-utterance visual scanning should require more time to complete before the onset of an utterance (Brown-Schmidt and Konopka, 2011). Thus we ask three main research questions:

1. How informative are speakers when referring to objects in simple and more complex visual scenes?
2. Do fixations to contrast objects prior to utterance onset predict informativeness?
3. Do underinformative utterances have shorter speech onset latencies than informative ones?

We hypothesise that: (1) speakers will be highly informative in this simple referential task, producing underinformative referring expressions rarely, especially in simple displays; (2) increased looks to the contrast object will result in informative referring expressions and decreased looks will result in underinformative ones; (3) underinformative utterances will have shorter onset latencies than informative ones. Based on high rates of between-speaker variability in informativeness found by Davies and Katsos (2010), we will also conduct an exploratory analysis of the role of individual differences in referring behaviour and hypothesise that there will be a distinctive linguistic-cognitive profile for underinformative vs. informative speakers.

## 2. Method

### 2.1 Design

The experiment used a 2 x 2 (contrast x display complexity) within-subjects design. Contrast was present or absent (two referents vs. one referent from the same noun category). Display complexity was 4- or 8-objects. Thus, for investigating the form of referring expressions in participants' production data (section 3.1), contrast and display complexity entered the analysis as independent variables. The dependent variable was utterance type (i.e. informativeness): underinformative, optimally informative, or overinformative.

The contrast variable was dropped from the analyses of eye movements (section 3.2) and speech onset times (section 3.3), because they focused on processing of the contrast object, which was of course absent in the contrast-absent condition. Thus, only contrast-present items were included. In addition, these analyses used informativeness (utterance type) as an independent variable, though only with two levels: underinformative and optimally informative (the latter coded as *informative* in all eye movement analyses). Overinformative trials were excluded due to their low frequency in the data. This resulted in what was basically a second 2 x 2 (display complexity x informativeness) design.

The dependent variables were: 1) Eye movements, i.e. a) *proportion of trials* in which speakers fixated the contrast object before producing an informative or an underinformative utterance, and b) *fixation time* to the contrast vs. the target object during four temporal regions (preview/pre-/during/post-utterance). *Fixation counts* to the contrast or target objects in the same temporal regions are also reported where relevant; 2) Speech onset time, as measured from the point at which the target object was highlighted to the onset of the participant's utterance 'click on the X'.

2.2 Participants

25 participants were recruited from the University of Leeds, UK. All were native speakers of British English, and all had normal or corrected-to-normal vision and hearing. Each participated voluntarily. One participant was excluded due to lack of attention throughout the testing session, leaving the remaining sample at n = 24, mean age 19 years (SD = 1.5), four males.

2.3 Materials

2.3.1 Referential communication task

The experiment consisted of 101 displays of everyday objects, grouped into semantically related sets, e.g. animals, food, household objects, clothes. Of these displays, 32 were experimental items, 64 were fillers and five formed the practice block. All objects were presented in grayscale and fit within square areas of interest measuring 300 x 300 pixels (4-object displays) and 235 x 235 pixels (8-object displays). As participants were seated 60 – 70cm from the monitor screen, the areas of interest surrounding each object spanned approximately 7° of visual angle for 4-object displays and 5.5° for 8-object displays.

For the experimental items, half of the displays contained four objects and half contained eight objects (see Figs. 1 and 2 for example displays), constituting simple and complex displays respectively. In addition, half of the experimental displays contained no contrast, with only one referent of each noun category (e.g. a ball, a doll, a teddy and a car, see left-hand panel of Fig. 1) and half contained a contrast display featuring two referents of the same noun category (e.g. a sweet, a large sausage, a small sausage and a sandwich, see right-hand panel of Fig. 1); one of these was the target and thus required modification for disambiguation. Target objects differed from their contrast mates on one of two scalar dimensions: size (large vs. small) or length (long vs. short). The small objects were created by resizing their larger counterparts. The short/long counterparts were created from separate images but were nevertheless highly visually similar. In the 4-object displays, the no-contrast items thus contained three distractor objects and the contrast items contained two distractors. In the 8-object displays, the no-contrast items contained seven distractor objects and the contrast items contained six distractors. These 32 critical items all appeared in four pseudorandomised lists, counterbalanced for target attribute and for block order, meaning that half the participants saw e.g. the small apple as the target while the other half saw the large apple as the target. No target object appeared more than once throughout the experiment, and the position of the target and the contrast object was rotated around each slot within the 4- and 8-object displays.

The 64 filler items were of four types: 16 x single object displays, 16 x two-object displays, 16 x 4-object and 16 x 8-object displays. In the 4- and 8-object displays, targets differed from contrast mates on one of two non-scalar dimensions: pattern (stripy vs. spotty) or openness (open vs. closed). The fillers were designed to mask the pattern inherent in the critical trials and to reduce predictability: in half of the filler items, a contrast set was present but neither of its members was the target.

The sequencing of each trial is depicted and explained in Fig. 3. The sequencing and timing is comparable to that used by Vanlangendonck et al. (2016). Stimuli were presented and eye movements were recorded using Tobii Studio version 3.2.3. The experiment was conducted using a Tobii X120 remote desk-mounted eye tracker sampling at 120Hz, a Dell 17” flat panel monitor with a content area of 1280 x 1024 pixels visible to the participant, and a Lenovo W540 laptop running the experimental software, visible to the experimenter. Participants’ utterances were recorded using a tabletop microphone, and they were video-recorded throughout using an external webcam running through Tobii Studio.

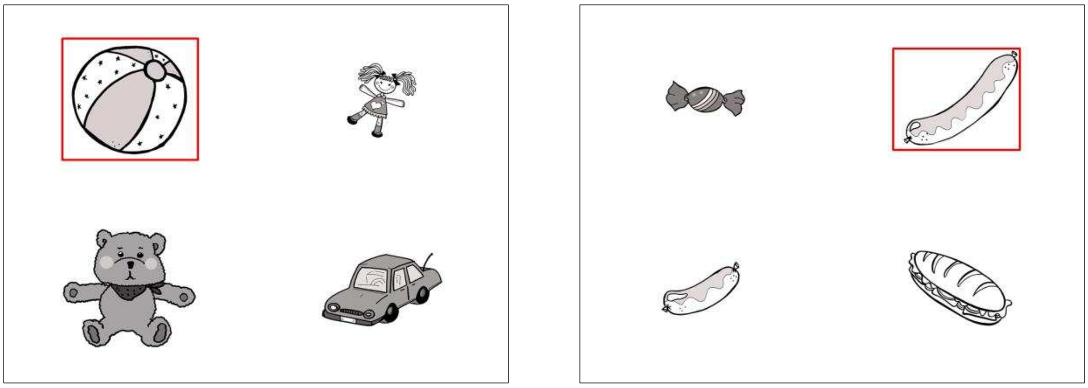

Figure 1. Example 4-object displays. Left hand panel shows a no-contrast item and right hand panel shows a contrast item. Target is highlighted in both panels in the same way as for the participant; the contrast object is the small sausage in the right-hand panel.

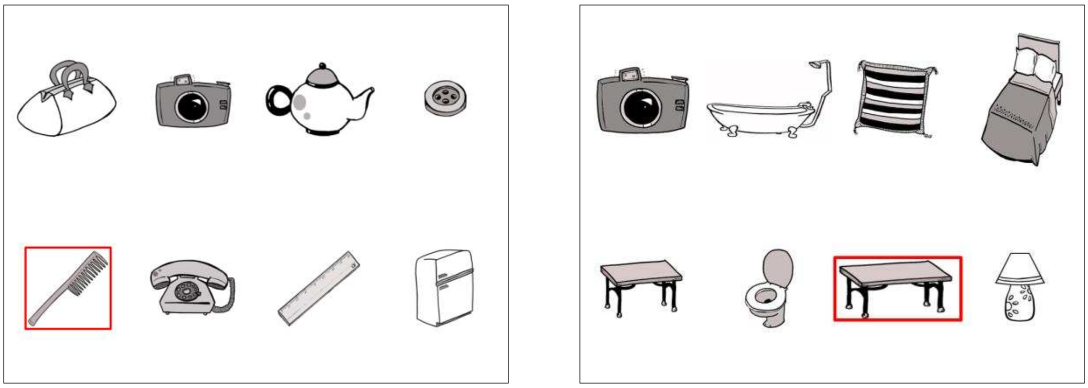

Figure 2. Example 8-object displays. Left hand panel shows a no-contrast item and right hand panel shows a contrast item. Target is highlighted in both panels; the contrast object is the shorter table in the right-hand panel.

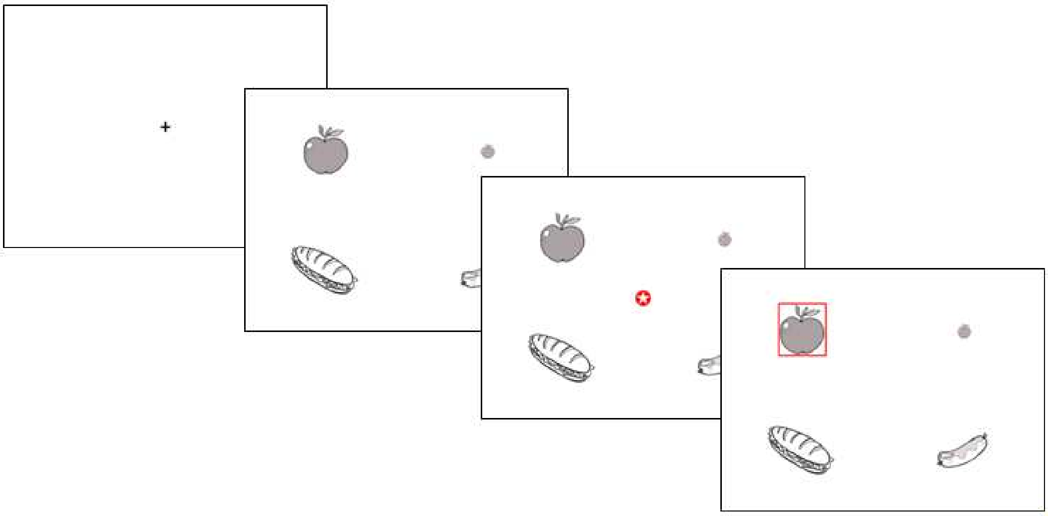

Figure 3. Trial sequence. 1) The fixation cross was presented for 1000 ms, followed by 2) a preview of the displays without the target highlight for 2000 ms (4-object displays; 3000 ms for 8-object displays). 3) A red fixation cross then appeared within the preview for a further 1000 ms. 4) The fixation cross disappeared and the target was highlighted with a red frame around the object. This final display remained visible for 4000 ms, during which time the participant produced their utterance using the form 'click on the X'.

2.3.2 Standardised Tests

As a further exploratory measure, five tests of linguistic and cognitive abilities were administered in order to correlate participants' profiles with their informativeness in the referential communication task. These profile measures also served as control data for a separate study on children's informativeness (Davies and Kreysa, in prep.). As an index of expressive language ability, the Expressive One Word Picture Vocabulary Test was used, normed for ages 2 - 80+ (Brownell, 2000). For receptive language ability, the British Picture Vocabulary Scale (BPVS-III) was used, normed for 3 - 16 year-olds, but recommended for use throughout the lifespan (Dunn, Styles and Sewell, 2009). For visual search efficiency, the visual search task from the PEBL battery was used (Mueller, 2014): this measures participants' speed and accuracy while they are searching for one or more letter targets defined by colour and shape in a field of distractors. To measure inhibition, the colour Stroop task recommended for evaluation of attentional filtering was used, also from the PEBL battery (Mueller, 2014). As a measure of perspective-taking ability within a discourse context, the Short Narrative subtest from the DELV-ST (Diagnostic Evaluation of Language Variation) Screening Test, recommended for use with 4 - 9 year-olds (Seymour, 2003). The study was approved by the Faculty ethics committee at the lead author's institution.

2.4 Procedure

The testing session took place in a purpose-designed lab in the Linguistics and Phonetics department at the University of Leeds. Participants were briefed on the content of the testing session and gave their informed consent. The order of tasks was the same for all participants, as follows (with approximate running times):

1. Referential communication task. Participants were seated in front of the eye tracker and monitor, with the experimenter seated nearby at the laptop. The two screens were not

mutually visible. A nine-point calibration was performed, then written experimental instructions were presented on screen. Participants were asked to tell the experimenter to click on the highlighted item. They were told that the experimenter could see the same objects that they could but that the objects were not in the same position on screen for him/her (to discourage reference using location), nor was the target highlighted. During the experiment, the experimenter issued a mouse click to signal that they had found the referent roughly one second after the offset of the participant's utterance, regardless of the form of referring expression used. No other feedback was given. The task was split into four blocks of equal length with voluntary breaks between (total 15 mins).

2. Expressive One Word Test, administered on hard copy according to the manual's instructions (10 mins).
3. British Picture Vocabulary Test, administered on hard copy according to the manual's instructions (10 mins).
4. Visual Search Test from the PEBL battery, run on a PC (10 mins).
5. The Short Narrative subtest from the Diagnostic Evaluation of Language Variation, administered on hard copy according to the manual's instructions (3 mins).
6. Stroop colour task, run on a PC (10 mins).

Participants were fully debriefed as to the aims of the experiment. The whole test session lasted approximately 75 minutes.

### 2.5 Data preparation

#### 2.5.1 Utterance coding

The utterances were transcribed and coded using the following system. If a referring expression contained minimally sufficient information for the addressee to uniquely identify it (i.e. with appropriate modification in the contrast-present condition) it was coded as informative. If it lacked such information (e.g. 'click on the apple' in the contrast-present condition) it was coded as underinformative. Since we were interested in the first attempt at referring, utterances which were initially underinformative but subsequently self-corrected to an informative form were coded as underinformative (e.g. 'click on the scissors - the small scissors'). 26 out of the 384 critical referring expressions (7%) were of this type. Referring expressions which contained more information than necessary for unique reference resolution (e.g. 'click on the long stripy sock' in a display with a long and a short sock, both stripy) were coded as overinformative. Utterances which referred to an incorrect target or were otherwise incomprehensible were coded as 'other' and excluded from subsequent analysis (two of the 384 critical utterances were of this type).

#### 2.5.2 Preparation of the eye tracking data

The *preview* temporal region was identified in the eye movement data as illustrated in Fig. 3 (screen 2), i.e. the period beginning with the display being revealed and ending with the red fixation cross appearing. Onsets and offsets of all critical utterances were calculated using the Sound Finder function in Audacity (Audacity Team, 2014), and then manually adjusted where required (e.g. where the function had falsely detected background noise as the speaker's voice). These exported timestamp labels were merged into the eye tracking data exports to provide utterance duration information. By cross-referencing utterance duration information with the timestamps for onsets and offsets of each visual stimulus, we defined three further temporal regions: before, during, and after the utterance. The *pre-utterance* temporal region was the period beginning with the target being highlighted and ending with the onset of the speaker's utterance (this region was equivalent to speech onset time); the *utterance* temporal region was equal to the duration of the utterance, and the *post-utterance* temporal region was the period from the offset of the utterance to the offset of the visual stimulus at the trial end.

Areas of interest (hereafter 'objects') in the displays were coded as Target, Contrast (if present), and Distractor. Fixation counts and total fixation durations to each object during each of the four temporal regions were then derived. Fixations which spanned multiple temporal regions in the pre-utterance, utterance, and post-utterance regions were allocated to all of the regions in which they occurred, as follows: for fixations which spanned two regions, half a fixation was allocated to each of the relevant regions. For fixations which spanned three regions, .33 of a fixation was allocated to each of the relevant regions. Fixation duration included individual fixations, gazes (i.e. two or more consecutive fixations to the same object), and refixations within one temporal region.

Only eye movements in the contrast condition (16 trials per participant) were analysed, since we were primarily interested in fixations to the contrast object, which was absent in the no-contrast condition. Four participants were wholly excluded from the eye tracking analysis since less than 50% of their samples recorded were usable, leaving the remaining sample at n = 20, mean age 18 years (SD = 4), four males. In addition, since some of the remaining critical trials contained invalid samples (i.e. samples rated 4-4 by Tobii Studio, indicating that neither eye had been found by the eye tracker), we used a data-driven approach to exclude any individual trials which contained less than 50% of the mean number of valid samples per trial, for each participant. As a result, 300 trials in the preview region and 301 in the region spanning pre-/utt-/post-utterance went forward for analysis.

Since the number of distractors varied across display types (three in 4-object 1-referent displays, two in 4-object 2-referent displays, seven in 8-object 1-referent displays, and six in 8-object 2-referent displays), total fixation values to distractor objects were corrected to ensure that they were comparable with values to the target and contrast objects. This correction was done by dividing fixation durations to distractor objects by the number of distractor items in each display type. Finally, the referential form coding (underinformative; informative) was merged with the eye tracking data.

2.5.3 Analyses

Most statistical analyses were performed using R (R Core Team, 2015), in particular the lme4 package (Bates, Maechler, Bolker and Walker, 2015). Unless otherwise mentioned, mixed effects analyses were conducted on the basis of initial maximal models, including random intercepts for both participants and items, and random slopes with all fixed factors. Fixed factors were either dummy-coded (binomial models) or centered around a mean of zero to minimize collinearity (log-ratios, fixation durations, and speech onset times). Models were fitted by maximum likelihood, with log-likelihood ratio tests ascertaining whether the interactions in the fixed-effects structure improved model fit for the maximal compared to simpler models. Where this was not the case, interactions were removed from both the fixed and the random parts of the models. For linear mixed effects models we report the coefficients, SE and t-values for all fixed effects and interactions in the final models. Only coefficients for which the absolute value of the t-statistic was greater than 2 are reported as significant.

## 3. Results

3.1 Referential communication task: Production data

All except one of the modified referring expressions took a prenominal adjective of the form 'click on the [adj](er) [noun]'. The only postnominally modified token was in reference to one of the filler items: 'click on the water bottle that's open'. This preference for prenominal modification is in line with Brown-Schmidt and Tanenhaus's (2006) findings for references to simple shapes.

In an analysis of all production data (contrast and no contrast conditions; 4- and 8-object displays, see table 1; selected data reproduced in Fig. 4), speakers were highly informative in their choice of referring expressions, with a mean rate of 88% of the total expressions produced being optimally informative. Because overinformativeness was extremely rare in the data (2% of all utterances), comparisons focus on rates of optimal informativeness vs. underinformativeness. A Wilcoxon signed-rank test shows that there was an effect of display complexity ($Z = -2.14$, $p < .05$, $r = .44$), with a larger number of optimal expressions produced for the simpler 4-object displays (mean 90%) than for the more complex 8-object displays (mean 86%). There was also an effect of contrast ($Z = -3.83$, $p < .001$, $r = .78$), with optimal expressions produced more often in the no-contrast condition (mean 97%) than in the contrast condition (mean 79%).

A Friedman's ANOVA backed up these main effects of display complexity and contrast on informativeness, $\chi^2(3) = 35.37$, $p <.001$. For interaction effects, post hoc analyses with Wilcoxon signed-rank tests were conducted with a Bonferroni correction applied, resulting in a significance threshold set at $p < .013$. Comparing between contrast conditions, mean rates of informativeness were significantly more frequent in the no-contrast conditions than in the contrast conditions (4 object: $Z = -3.44$, $p = .001$, $r = .70$; 8 objects: $Z = -3.67$, $p < .001$, $r = .75$). Comparing between display complexity conditions (see Fig. 4), informativeness in the 4-object contrast condition was significantly more frequent than in the 8-object contrast condition ($Z = -3.05$, $p = .002$, $r = .62$). These comparisons suggest that speakers are less likely to be informative in contrast displays, especially for 8-object displays. Note however, that this contrast effect is driven by the zero rate of underinformativeness in the no-contrast condition: to be underinformative in the absence of contrast would require participants to underspecify a bare noun, e.g. by saying 'click on that'.

| | 4-object displays | 8-object displays | Total displays (4 and 8 combined) |
|---|---|---|---|
| **No-contrast condition: mean % (SD)** | | | |
| Underinformative | 0 (0) | 0 (0) | 0 (0) |
| Optimally informative | 96 (7) | 97 (6) | 97 (4) |
| Overinformative | 3 (7) | 3 (6) | 3 (5) |
| **Contrast condition: mean % (SD)** | | | |
| Underinformative | 15 (22) | 24 (22) | 19 (21) |
| Optimally informative | 84 (21) | 75 (22) | 79 (21) |
| Overinformative | 1 (4) | 2 (5) | 2 (4) |
| **Total (combined contrast and no-contrast conditions): mean % (SD)** | | | |
| Underinformative | 7 (11) | 12 (11) | 9 (10) |
| Optimally informative | 90 (12) | 86 (11) | 88 (10) |
| Overinformative | 2 (4) | 2 (4) | 2 (3) |

Table 1. Mean rates of informativeness as a percentage of all expressions produced.

Fig. 4 presents production results for the contrast condition only due to its importance in the eye movement analysis in section 3.2. Speakers were largely informative in their choice of referring expression in the contrast condition, with 79% of all expressions being optimally informative.

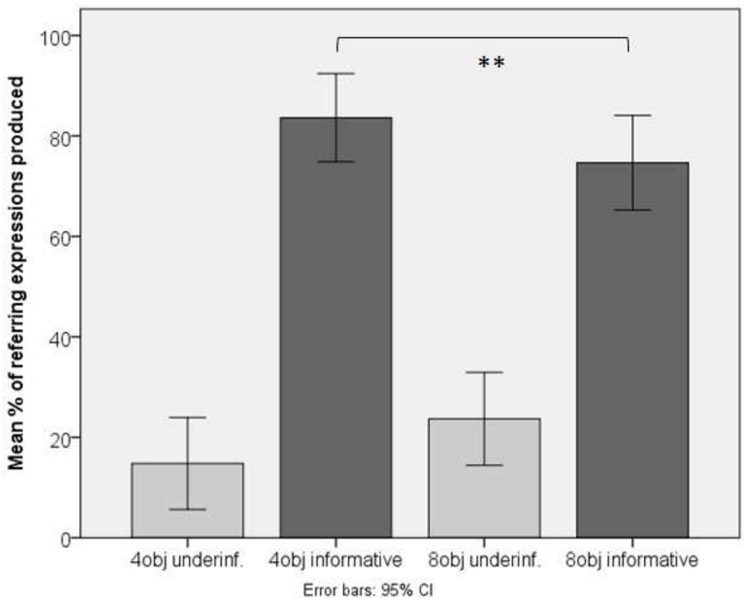


Figure 4. Mean rates of informativeness of referring expressions as a percentage of all expressions produced; contrast items only, ** $p < .005$.

Summarising the production results, speakers were largely informative when referring to objects for their addressee, both overall (88%) and in the contrast condition where they had to produce a disambiguating adjective (79%). They produced underinformative expressions 19% of the time in the contrast condition, while overinformative expressions were rare across conditions. This confirms our first hypothesis, i.e. that speakers would be highly informative in this task. Regarding display complexity, speakers were more likely to produce underinformative referring expressions for more complex displays than for simpler displays.

### 3.2 Eye movement data

#### 3.2.1 Contrast fixations during the preview and pre-utterance temporal regions

To investigate the relationship between fixation of the contrast object and speaker informativeness, we looked at the proportion of trials in which speakers fixated the contrast object before producing an informative or an underinformative utterance. Trials that were invalid in one or both of these regions were excluded, leaving 92% of the original dataset.

Trials were categorised as showing one of four fixation patterns: a contrast fixation in i) the preview region alone, ii) the pre-utterance region alone, iii) both regions, or iv) neither region. The mean proportions of underinformative utterances by fixation pattern were calculated (e.g. 38% of all trials involving a contrast fixation in neither the preview nor the pre-utterance region were underinformative) and are shown in Fig. 5.

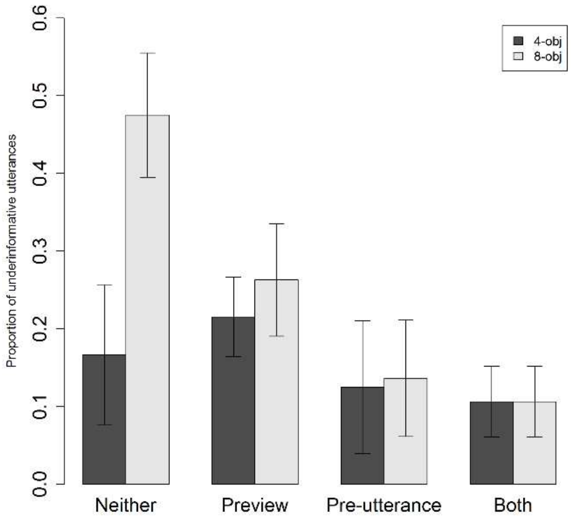


Figure 5. Mean proportions of underinformative utterances following contrast fixation patterns across two temporal regions. Proportions and SDs for 4- and 8-object displays combined are: 38% (49) for fixations in Neither region, 23% (42) for Preview only, 13% (34) for Pre-utterance only, and 11% (31) Both. Error bars show +/- 1 SE.

To analyse the role of contrast fixations in the informativeness of the subsequent utterance, we used generalised linear mixed effects models assuming a binomial distribution. An initial analysis predicted the occurrence of an underinformative utterance based on the temporal region(s, if any) in which the contrast object was fixated, across the two display complexity conditions. Subsequently, we split the dataset by display complexity, and ran separate analyses for 4- vs. 8-object displays. In all cases, the four contrast fixation patterns (both, neither, pre-utterance, preview) were dummy-coded with “both” as the reference level. The maximal model thus included contrast fixation pattern as a fixed effect, participants and items as random effects, and a random slope for contrast fixation pattern by item (i.e. *informativeness ~ contrast fixation + (1| ppt) + (1+contrast fixation | item).* Convergence was achieved using the bobyqa optimiser. Overall, as depicted in Fig. 5, speakers were more likely to be underinformative if they never fixated the contrast object than if they fixated it in both the preview and the pre-utterance regions (estimate = 2.123, *SE* = .865, *p* = .014). There was no significant difference between fixating the contrast in the pre-utterance region only and fixating it in both regions (estimate = .892, *SE* = 1.003, *p* = .374), and fixating the contrast in the preview region made speakers only marginally more likely to be underinformative than fixating it in both regions (estimate = 1.281, *SE* = .685, *p* = .061). In other words, speakers were most likely to produce an underinformative expression if they did not previously fixate the contrast object in either temporal region, and least likely to do so if they fixated it in both. Fixating the contrast in the preview region alone slightly reduced the likelihood of producing an underinformative expression, whereas fixating the contrast in the pre-utterance region was equivalent to fixating it in both regions. Although the difference is small, this suggests that fixating the contrast just before articulation is more helpful for producing an informative expression than earlier fixations before the target has been identified.

Splitting the data by display complexity (using the same model structure as above) reveals that these effects are driven by the complex displays. Among the 4-object trials, there were no significant effects of contrast fixation pattern on informativeness (all *p*s > .1): rates of underinformativeness were similarly low in all conditions (neither: *M* = 17%, *SD* = 38; preview: *M* = 22%, *SD* = 41; pre-utterance: *M* = 12%, *SD* = 34; both: *M* = 11%, *SD* = 31). In contrast, among the 8-object trials, speakers were substantially more likely to be underinformative if they never fixated the contrast object (*M* = 47% of trials were underinformative, *SD* = 51) than if they fixated it in both preview and pre-utterance regions (*M* = 11% of trials were underinformative, *SD* = 31; *p* = .002). Again, there was no significant difference in informativeness between fixating the contrast in the pre-utterance region only and fixating it in both regions (*M* = 14% of trials were underinformative, *SD* = 35, *p* > .6). The same was true for fixating the contrast in the preview region only (*M* = 26% of trials were underinformative, *SD* = 45, *p* > .5). In other words, speakers were most likely to produce an underinformative expression if they did not previously fixate the contrast object in either temporal region. That is, for complex displays, informativeness depended strongly on the occurrence of at least one contrast fixation at any time before utterance onset. However, even for complex displays, the contrast fixation was by no means essential: informative utterances were produced in 53% of trials in which no contrast fixation had occurred.

An additional analysis of fixation duration to the contrast object corroborated the binary findings above. Linear mixed effects models investigated the influence of display complexity and informativeness on fixation duration to the contrast object during the preview and pre-utterance temporal regions. Since there were 58 trials in which speakers did not fixate the contrast at all in these regions, we excluded those trials from this analysis. Seven outlying trials with fixation durations of >1250ms were also excluded, leaving 77% of the prepared dataset. The maximal model included the two fixed factors (display complexity and informativeness), their interaction, random intercepts for participants and items, and random slopes for the interaction of display complexity and informativeness by participants and by items: *fixtime to contrast ~ display complexity * informativeness + (1+ display complexity * informativeness | ppt) + (1+ display complexity * informativeness |item)*.

As Fig. 6 shows, during the preview and pre-utterance regions combined, speakers showed a tendency to fixate the contrast object longer before producing an informative utterance (*M* = 432 ms, *SD* = 235) than when producing an underinformative utterance (*M* = 362 ms, *SD* = 253; informativeness coefficient = -90.3, *SE* = 56.2, *t* = -1.60). Thus, longer looks to the contrast object before speaking seemed to be associated with informativeness. The complexity of the display did not affect fixation duration either on its own (*t* = .26) nor in its interaction with informativeness (*t* = .54).

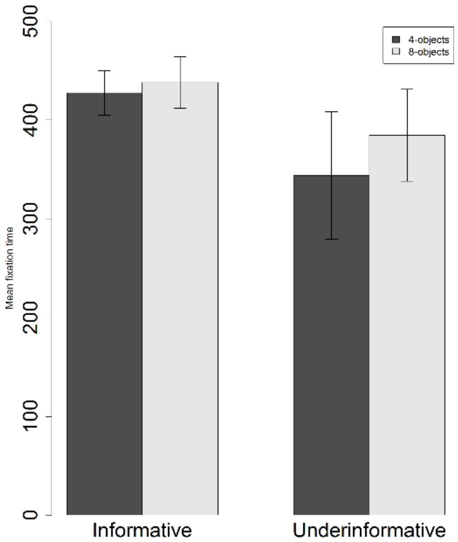


Figure 6. Mean total fixation duration to the contrast object during the preview and pre-utterance regions, by informativeness and display complexity. Error bars show +/- 1 SE.

3.2.2 Fixation time to contrast vs. target object, by informativeness

Although the previous section revealed clearly that contrast fixations are helpful in producing an informative utterance, it is interesting to observe that a substantial number of trials did not contain a single contrast fixation in either the preview or the pre-utterance temporal regions. This raises the question of where speakers *were* looking while planning their utterance. Here we show that instead of fixating the contrast object in the period between learning what the target was and articulating their utterance, speakers tended to spend most of the time fixating on the target object. This tendency continued throughout the entire utterance, though it was most pronounced in the pre-utterance phase.

In order to analyse the relative preference for the contrast vs. the target object, we compared mean log probability ratios for fixations to the target object relative to the contrast object over time, for fully informative and for underinformative utterances *(ln(P(target)/P(contrast))*. A log probability ratio of zero would indicate equal attention to the target and the contrast object; a positive score implies that the target was fixated more than the contrast object, and vice versa for a negative score. Because log-ratio analyses rely on aggregated fixation durations and in order to avoid substantial occurrences of trials with zero values due to low overall rates of contrast object fixations, we fitted separate linear models averaged over participants and items. In both cases, the data was first aggregated to reflect the mean fixation durations per AOI, temporal region and informativeness (utterance type) for each participant or item. Log-ratios were then computed for fixations to the target over the contrast object in each of the four temporal regions, avoiding division by zero by replacing empty cells with 0.01, i.e.

a very small number. This transformation resulted in minimal (i.e. inconsequential) changes regarding the mean and median values.

The initial models included two fixed factors, utterance informativeness and temporal region, as well as their interaction, random intercepts for participants or items, and a random slope for temporal region. The interaction reduced model fit for the maximal compared to simpler models both by participants and by items, so was omitted from the final model *logratio_contrast ~ informativeness + temporal_region + (1+ temporal_region | ppt[/item]*).

As Fig. 7 shows, in all temporal regions speakers substantially preferred to fixate the target object longer than the contrast object (significant intercept by participants = 2.635, *SE* = .312, *t* = 8.45; by items = 1.898, *SE* = .443, *t* = 4.29), regardless of the utterance type they were producing. This preference – which conversely implies shorter durations for fixating the contrast object – was particularly pronounced when they produced an underinformative utterance (*M* (target preference) = 4.0, *SD* = 5.27 for underinformative vs. *M* = 1.3, *SD* = 1.61 for fully informative utterances). The effect of informativeness was fully significant by participants (*t* = 4.42) and marginally significant by items (*t* = 1.96), while the effect of temporal region was fully significant both by participants (*t* = 2.76) and by items (*t* = 2.20). This pattern of results is reflected clearly in Fig. 7: the strongest target preference was found in the pre-utterance phase (*M* = 4.39, *SD* = 4.29), the weakest target preference (unsurprisingly) in the preview region (*M* = 0.21, *SD* = 3.71), when speakers did not yet know which object would turn out to be the target. Importantly, the strongest tendency to fixate the target (and thus to ignore the contrast object) occurred in the pre-utterance phase of those utterances that would subsequently be underinformative (*M* = 7.19, *SD* = 4.59), in contrast to those that would prove to be informative (*M* = 1.6, *SD* = 0.58). The same pattern of results was found for an analysis of mean log probability ratios for number of fixations (cf. fixation duration) on the target object relative to the contrast object for fully informative and for underinformative utterances.

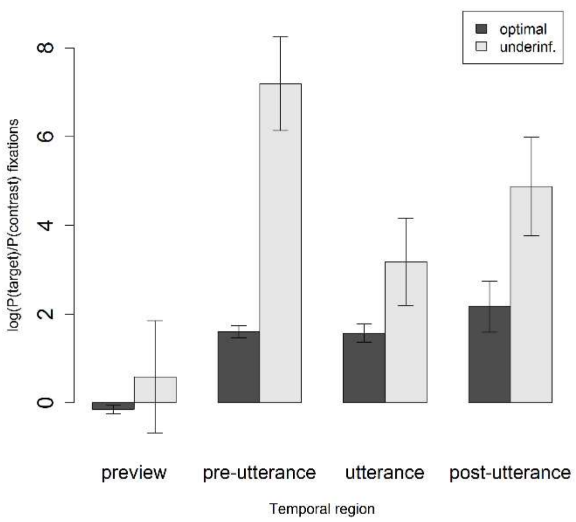


Figure 7. Visual preference for the target over the contrast object in all temporal regions as a function of informativeness. Error bars show +/- 1 SE.

In combination with the analysis in section 3.2.1, these results confirm our second hypothesis. That is, increased looks to the contrast object resulted in informative referring expressions and decreased looks resulted in underinformative ones. We also revealed that the target object is preferred overall, regardless of informativeness.

### 3.2.3 Contrast fixation duration during articulation

Although we ascertained in section 3.2.1 that speakers were more likely to be informative if they fixated the contrast object before starting to speak, it is possible that later contrast fixations might also have influenced their choice of referring expression. Brown-Schmidt and Tanenhaus (2006) found that contrast fixations after the onset of the referring expression resulted in postnominally modified (or repaired) referring expressions (see also Brown-Schmidt and Konopka, 2011). Although our data did not contain any postnominal modification, we analysed post-onset contrast fixations to comprehensively analyse the timecourse of the relationship between contrast fixations and informativeness. This analysis is also motivated by Pechmann's (1989) argument that articulation begins before visual scanning is complete. Under this incremental account, it would be perfectly possible for fixations made during the utterance to boost informativeness.

Linear mixed effects models were used to investigate the influence of display complexity and informativeness on fixation duration to the contrast object during the utterance temporal region. Since there were 137 trials in which speakers did not fixate the contrast at all during the utterance temporal region, we excluded those trials from this analysis. Two outlying trials with fixation durations of >1250ms were also excluded, leaving 54% of the prepared dataset for analysis. The initial maximal model included the two fixed factors display complexity and informativeness, their interaction, random intercepts for participants and items, a random slope for display complexity by participants, as well as a random slope for the interaction of display complexity and informativeness by items only. Because this model did not converge, we removed the interaction term from the by-trial random slope (but kept it as a main effect by trial), thus obtaining the final model: *fixtime to contrast ~ display complexity * informativeness + (1+ display complexity * informativeness | ppt) + (1+ display complexity + informativeness |item)*.

As Fig. 8 shows, during articulation, speakers fixated the contrast object longer when producing an underinformative utterance (*M* = 417 ms, *SD* = 279) than when producing an informative utterance (*M* = 315 ms, *SD* = 198; informativeness coefficient = 164.5, *SE* = 67.8, *t* = 2.43). Thus, longer looks to the contrast object while speaking were associated with underinformativeness. Although the complexity of the display on its own did not affect fixation duration (*t* = -.99), it approached significance in its interaction with informativeness (interaction coefficient = -215.8, *SE* = 116.2, *t* = -1.86), such that when producing an underinformative utterance, speakers looked at the contrast longer for 4-object displays (*M* = 480 ms, *SD* = 333) than for 8-object displays (*M* = 369 ms, *SD* = 226). No such difference was found for informative utterances.

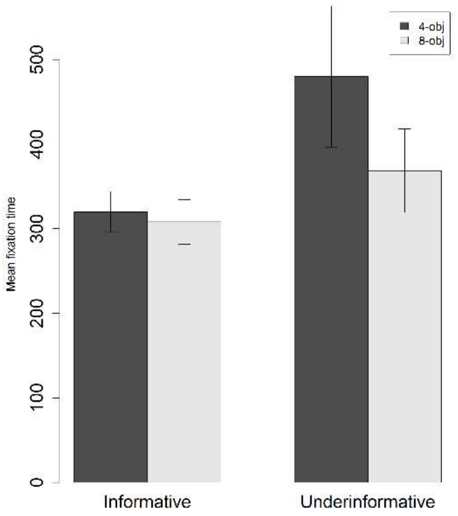


Figure 8. Mean fixation duration to the contrast object during the utterance temporal region, by informativeness and display complexity (fixated trials only). Error bars show +/- 1 SE.

One explanation for the longer contrast fixation durations during underinformative utterances in simple displays could be that they are due to self-corrections. To explore this possibility, we first checked whether the 22 self-corrected trials actually involved a contrast fixation in the utterance region. This was the case for 17 of these trials. Descriptively, these self-corrected trials (all underinformative by definition) had higher mean contrast fixation durations ($M$ = 507 ms, $SD$ = 290) than the 11 contrast-fixated underinformative trials without self-corrections ($M$ = 400 ms, $SD$ = 213), but the counts were too low for statistical analysis.

To summarise the analysis of fixations during the utterance region, speakers fixated the contrast object for longer while producing an underinformative utterance than when referring informatively. This difference was more marked for simple displays, and may have been mediated by self-corrections. Thus, looking longer at the contrast object after the onset of an utterance was not effective for encoding information about its distinctive features into a referring expression. Interestingly, this pattern differs from the relationship between contrast fixations and informativeness found in the preview and pre-utterance temporal regions. The utterance region analysis thus complements the findings related to our second hypothesis: although the contrast object was fixated more prior to informative referring expressions than before underinformative ones, later contrast fixations did not boost informativeness in the same way.

### 3.3 Speech onset time

We analysed speech onset time (SOT) in order to test the third hypothesis, i.e. that underinformative referring expressions would have shorter onset latencies than informative ones. Speech onset time was defined as the period between the target being highlighted and the onset of the speaker's utterance, which was equivalent to the pre-utterance region in the eye movement analyses. The analysis incorporated all trials from the contrast condition which were either underinformative or optimally informative. Two outlying data points above 2500 ms were excluded. See Table 2 for results.

| | 4-object displays | 8-object displays | Total displays (4 and 8 combined) |
|---|---|---|---|
| Informative | 1074 (238) | 1142 (245) | 1106 (244) |
| Underinformative | 1064 (297) | 1116 (282) | 1095 (287) |
| Total | 1072 (248) | 1136 (254) | 1103 (253) |

Table 2. Mean (SD) speech onset times in ms by display complexity and utterance informativeness.

Before splitting by informativeness and display complexity, overall speech onset times replicated those found by Rabagliati and Robertson (2016). A linear mixed effects model was used to investigate the influence of display complexity and informativeness on SOT. The final model *(SOT ~ display complexity * informativeness + (1+ display complexity| ppt) + (1+ display complexity + informativeness |item))* revealed that speech onset was earlier for underinformative utterances than for informative utterances (informativeness coefficient = 87.8, *SE* = 31.1, *t* = 2.82). Speech onset was also earlier for less complex displays (complexity coefficient = 60.5, *SE* = 22.4, *t* = 2.70). The interaction between informativeness and display complexity was not significant (*t* = -.44).

To investigate the relationship between SOT, contrast fixations and informativeness, we ran partial correlations between SOT and the occurrence of contrast fixations (binary-coded) in the pre-utterance region, between SOT and level of informativeness produced, and between occurrence of contrast fixations (binary-coded) in the pre-utterance region and level of informativeness produced, controlling for display complexity in all cases. There was a positive correlation between SOT and occurrence of contrast fixations (*r* = .25, *p* = .001), no association between SOT and level of informativeness (*r* = .03, ns), and a positive correlation between contrast fixations and level of informativeness (*r* = .23, *p* = .001; in line with findings presented in section 3.2.1).

To summarise the analysis of speech onset times; participants began speaking slightly but significantly earlier when producing underinformative utterances, and also when referring to objects in simpler displays. This confirms our third hypothesis, suggesting that later speech onsets may allow time for extra fixations, which in turn may boost informativeness. Partial correlations revealed that longer speech onset times are associated with speakers fixating the contrast object during the pre-utterance region, but are not in themselves directly linked to increased informativeness.

## 3.4 Relationships between performance on standardised tests and patterns of informativeness in production.

### 3.4.1 Scoring of the test battery

Scores were calculated using the test manuals and the PEBL software. For both the visual search and the Stroop task we excluded any RTs <200 ms. For the Stroop, the upper cut-off was limited by the software to 2000ms and for the visual search we excluded responses of >5000 ms (2 SDs above the mean of the participant with the longest mean RT). Descriptive statistics can be found in Table 3.

| | Mean | SD | Minimum | Maximum | Range | Possible scores |
|---|---|---|---|---|---|---|
| EOWT (raw score) | 145.46 | 14.08 | 122 | 174 | 52 | 0 - 190 |
| BPVS (raw score) | 161.25 | 4.11 | 151 | 167 | 16 | 0 - 168 |
| DELV (raw score) | 5.88 | 1.19 | 3 | 7 | 4 | 0 - 7 |
| Visual search accuracy | 179.00 | 1.41 | 175 | 180 | 5 | 0 - 180 |
| Visual search mean RT (correct trials only) | 1361 | 266 | 979 | 1878 | 899 | cutoff at 5000ms |
| Stroop accuracy (incongruent trials only) | 52.25 | 1.87 | 50 | 56 | 6 | 0 - 56 |
| Stroop mean RT (incongruent trials only) | 830 | 125 | 649 | 1101 | 452 | timeout at 2000ms |

Table 3. Performance on the standardised tests (n = 24). Minimum, maximum and range refer to the scores attained by our participants; possible scores are defined by the construction of the tests.

3.4.2 Correlational analyses

A Pearson correlation coefficient was computed to assess the relationship between informativeness of referring expressions and performance on the standardised tests. No significant correlations were found between rates of underinformativeness in the contrast condition and performance on any of the standardised measures of language ability, perspective-taking or narrative abilities, visual search, or inhibition, (all *p*s > .1; all *rs* < .2). This lack of significant associations may have been due to low power (recall the mean rate of 19% underinformativeness) and the limited range of the scores (see Table 3). Individual differences were not investigated further but are mentioned briefly in the discussion below.

## 4. Discussion

This study examined speakers' eye movements alongside the informativeness of their referring expressions in order to explore associations between the two types of behaviour. Speakers produced fully informative referring expressions in the majority of their utterances. The analyses of fixation patterns revealed that speakers were more likely to be informative if they had fixated the contrast object during multiple temporal regions (Fig. 5) and for longer (Fig. 6) before starting to speak. However, such fixations were not essential for producing a fully informative utterance. This boosting effect of contrast fixation on informativeness was limited to the preview/pre-utterance temporal regions. Fixating the contrast after speech had begun was not associated with greater informativeness; longer fixation durations during speech were linked to greater underinformativeness (Fig. 8). In all temporal regions, speakers substantially preferred to fixate the target rather than the contrast object, regardless of the level of informativeness they produced (Fig. 7). Finally, speech onset times were shorter for underinformative utterances (Table 2). Taken together, our results suggest that fixating the contrast object while planning speech (which may result in a delay to the onset of an utterance) promotes fully informative referring expressions.

In line with our first hypothesis and with previous work on adult reference (Brown-Schmidt and Tanenhaus, 2006; Davies and Katsos, 2010; Heller and Chambers, 2014; Nadig and Sedivy, 2002; Rabagliati and Robertson, 2016; Vanlangendonck et al., 2016, i.a.), speakers were highly informative in this referential task, producing underinformative referring expressions rarely, especially in simple displays. They were able to acknowledge the presence or absence of a contrast object and modulate their referential choices accordingly. However, they provided underinformative utterances around 20% of the time by omitting the discriminating adjective. Although communicatively infelicitous, this referential variation allowed us to investigate how pre-utterance processing differed across levels of informativeness using eye tracking and speech onset analyses.

Speakers fixated the contrast object more frequently before informative referring expressions than before underinformative ones, confirming our second hypothesis. However, this pattern of results was complicated by the fact that in 62% of utterances, speakers were informative despite *not* having directly fixated the contrast object in either the preview or pre-utterance temporal region, closely replicating Brown-Schmidt and Tanenhaus's rate of 68% (2006, exp. 1). Thus, it appears that contrast fixations are helpful but not essential for full informativeness. Instead, speakers may be able to use information gleaned from extrafoveal vision to produce informative utterances. If this is the case, and assuming that larger contrast objects are easier to identify extrafoveally than smaller ones, trials containing large or long contrast objects should be more likely to result in informative utterances than those containing smaller or shorter ones. An additional analysis showed this to indeed be the case: of the 36 trials in which the utterance was informative and the contrast was not directly fixated in the preview and pre-utterance regions, 23 featured a big or long contrast object while only 13 featured a small or short contrast object. Further, we found that informative utterances without a contrast fixation were more common in simple displays (83% of all simple displays were informative despite no prior contrast fixations) than in more complex ones (53% of all complex displays), suggesting that sparser displays facilitate extrafoveal processing. However, it should be noted that modifier use in the unfixated contrast trials (62%) was starkly different to modifier use in no-contrast trials (3%), suggesting that participants were not *ignoring* the presence of the contrast object, but instead were able to use it - perhaps via extrafoveal processing - to produce informative expressions.

Although it is an intuitively plausible assumption that speakers can only accurately describe a visual scene in front of them if they have sufficiently processed the task-relevant aspects of this scene, our findings align with a growing number of studies using diverse tasks, displays and stimuli which suggest that comprehensive fixation of all such aspects is not essential for accurate and felicitous referring. This is the case especially for informationally lighter referential choices, with speakers looking less frequently and for shorter durations at given rather than new objects. In repeated reference, for example, speakers take a second look to check the relevant object or its properties only half of the time when producing an anaphoric pronoun (van der Meulen et al., 2001). More radically, in a study that required speakers to look continuously at a fixation point at the centre of a scene, speakers were found to be perfectly capable of describing depicted events (Griffin 2004:231). Scenes can be described with a high level of accuracy even after very brief presentation times, which effectively prevent fixations to all referents (Dobel, Gumnior, Bölte and Zwitserlood, 2007). These findings are complemented by several studies which find that object identities can be processed extrafoveally, that visual attention can shift to an object before it has been directly fixated, and that speakers can distribute their attention over multiple objects (Meyer, Ouellet and Häcker, 2008; Meyer, Belke, Telling and Humphreys, 2007; Morgan and Meyer, 2005; Schotter, Jia, Ferreira and Rayner, 2014). Taken together, these studies provide mounting evidence suggesting that it is not always necessary to look before speaking.

We have extended work showing that a target object is not necessarily fixated before a speaker refers to it and using novel displays, have shown a similar non-dependency for contrast objects. Instead, speakers have a strong tendency to fixate the target object – the one that was actually being described – over any other object and across temporal regions. This is predicted by the goal-based account of eye movements (Salverda et al., 2011) discussed in the introduction, which describes a mapping of language and eye movements such that the locations most relevant to the (linguistic) task at hand are also the ones most likely to be fixated. In line with this account, preference for the target is due to the fact that it is the object concurrently being referred to ('click on the x'). Indeed, even when contrast fixations *were* made, minimal looks were sufficient for full informativeness. Contrast fixations in the pre-utterance region, preceding informative utterances, were largely one-off and short (fixation count $M$ = 1.1, $SD$ = .6; fixation duration $M$ = 267ms, $SD$ = 128, excluding zero-fixation trials). Thus, these results suggest that visual scanning of unmentioned but relevant objects immediately before articulation is not the main factor in informative verbal behaviour, contrary to Pechmann's (1989) claim that incomplete scanning is to blame for failures in informativeness.

The boost to informativeness found as a result of preview and pre-utterance contrast fixations was not found in the utterance temporal region. In contrast with Brown-Schmidt and Tanenhaus's (2006) data on late looks being linked to postnominal modification, contrast fixations during speech did not lead to greater informativeness in our data. In fact, our speakers fixated the contrast for longer in simple displays during the articulation of *underinformative* compared to informative referring expressions. It is not clear whether this extended fixation was a cause of underinformativeness (i.e. despite being longer, these fixations came too late for encoding), or an effect (i.e. speakers may have been more aware of their inadequate referring expression while looking at simple displays, as conveyed by increased attention to the source of key discriminatory information in this condition). Indeed, fixations during the utterance region may in fact suggest greater self-monitoring behaviour during underinformative referring (cf. Rabagliati and Robertson, 2016) although it did not ultimately lead to self-correction. Future studies using finer-grained methods may be able to identify the point at which additional fixations no longer usefully recruit contrast information for use in reference production.

As predicted by our third hypothesis, underinformative utterances had slightly but significantly shorter speech onset times than informative ones, suggesting that the infelicitous utterances may be articulated before all relevant visual information had been acquired and sufficiently processed. This explanation is tentatively supported by the mean fixation counts and durations in the pre-utterance temporal region (underinformative fixation count: $M$ = 0.25, $SD$ = 0.5, duration: $M$ = 57 ms, $SD$ = 133; informative fixation count: $M$ = 0.6, $SD$ = 0.7, duration: $M$ = 137 ms, $SD$ = 160, including zero-fixation trials). The longer speech onset times for informative utterances may reflect time spent on contrast fixations or on planning the more complex informative referring expressions, or both. Future work would benefit from using methods enabling the two processes to be examined separately. The shorter speech onset times we observed for underinformative utterances may also be linked to variable onset times for different types of contrast, i.e. shorter onset times for colour (absolute) contrasts than size (scalar) contrasts (Brown-Schmidt and Konopka, 2011). Both our findings and those by Brown-Schmidt and Konopka support a serial process of speech conceptualisation, formulation, and articulation (Levelt, 1989), in which extensive pre-utterance visual scanning should require more time to complete before the onset of an utterance (contra accounts assuming that scanning and articulation occur in parallel, e.g. Pechmann, 1989).

Rather than complete visual scanning, we suggest that a more important factor driving speakers' behaviour in this kind of task is the cooperative goal of providing sufficient information for addressees to identify the target unambiguously. This Gricean drive is one of the fundamental assumptions of communication and has been documented in a large body of experimental pragmatics literature (see Noveck and Reboul, 2008 for a review). Our study adds to this evidence by documenting high levels of informativeness even in the absence of feedback and under time constraints. The nature of the task - helping a real, physically co-present addressee to fulfil an objective - may have added to the compulsion to provide felicitous referring expressions.

While pragmatic cooperativeness is a powerful constraint on referential choice, certain methodological aspects of our experiment should be acknowledged. For example, the high degree of visual similarity between target and contrast items in both size and length dimensions may simplify the task of identifying a contrasting same-category object without a direct fixation[1]. Given the evidence of extrafoveal processing cited above (e.g., Meyer, Ouellet and Häcker, 2008; Morgan and Meyer, 2005), it is quite plausible that the contrast object was processed extrafoveally during one or more of the analysed temporal regions, so did not manifest in overt fixations by the speaker. Indeed, the correspondence between fixations and visual attention is far from straightforward, and certain forms of visual displays may allow viewers to use a broader attentional focus and attend to several objects more or less in parallel (Cave and Bichot, 1999). Further research using multiple contrast sets, reduced preview opportunities or gaze contingent paradigms may shed light on these potential explanations for our findings as well as clarifying the timecourse of integrating visual information into the formulation of a felicitous description. In particular, display complexity, speech planning differences in the relevant object dimensions (e.g. relative vs. absolute distinctive features), and current discourse goals may all influence this process.

Although our task was reasonably interactive, we acknowledge that our stimulus displays were artificial. In the real world, objects don't appear mere seconds before speakers refer to them, they are situated within a rich visual context, and the decision about what to refer to is not usually cued by an external source. In addition, even our complex displays were much simpler than most real-world scenes. Nonetheless, our results showed that compared to even simpler 4-object arrays, the 8-object displays compromised informativeness, and presented a higher risk of underinformativeness when the contrast object was not fixated before speaking. It seems plausible that adding further complexity might amplify these findings. Equally plausibly, the importance of fixating contrast objects may have been somewhat masked by our non-naturalistic displays relative to real-world processing. Investigating the relationship between visual scanning and informative speech production using more naturalistic scenes (such as those used by Griffin, 2004 and Andersson, Ferreira and Henderson, 2011) would be welcomed.

Although contrast fixations were less predictive of informativeness than originally hypothesised, they did at least have a measurable effect, unlike factors relating to individual differences between speakers: participants' referential informativeness did not correlate significantly with their levels of

[1] We thank an anonymous reviewer for pointing out this possibility.

language ability, visual search efficiency, inhibition, or narrative ability (though as mentioned above, this lack of relationship may have been due to low rates of underinformativeness overall, plus the restricted range of scores on the standardised tests). The tendency to underinform was surprisingly consistent both within- and between-participants. This contrasts with data from Davies and Katsos (2010, exp. 1), where rates of underinformativeness were skewed by a subset of participants who underspecified more than 80% of their referring expressions, resulting in an overall mean rate of 27% underinformativeness. This disparity between studies may be due to differences in the methodology employed, e.g. a real vs. depicted addressee, and a predetermined vs. self-paced progression of the stimuli. Future research on the role of individual differences in referential informativeness may benefit from tasks eliciting higher rates of underinformativeness, e.g. using a dual-task paradigm to increase cognitive load.

In future work we would also like to examine the role of eye movements on processes leading to overinformativeness. Although our original plan was to analyse all three levels of informativeness in the taxonomy, only 2% of utterances in our dataset were overinformative, leading us to drop this level of comparison. Paradigms eliciting higher rates of overinformativeness, e.g. including two targets rather than a single target to find in a display (Koolen et al., 2011), or when a speaker's discourse goals involve teaching an addressee about an object (Arts, 2004), or when the precision of a description is paramount (Arts, Maes, Noordman and Jansen, 2011b) could be useful in investigating the link between contrast fixations and overinformativeness.

To sum up, this study has extended research investigating the chain of events leading to the production of expressions referring to objects in the visual world. It benefits from insights from theoretical work on reference, production studies looking at the relationship between eye movements and event descriptions, and the large body of research and methods investigating comprehension in the visual world. Working across these areas, we have provided useful evidence on the nature of the relationship between gaze and reference: fixating a contrast object immediately before speaking *facilitates* the production of an informative referring expression, but is not absolutely required. The cooperative speaker has a pragmatic drive to be informative and can use information gleaned a number of sources (direct fixation, extrafoveal processing, previous exposure) in order to provide their addressee with a referring expression at exactly the level of specification needed to identify the target object.

While few studies have looked at the relationship between contrast fixations and informativeness, our work makes a novel contribution in its analysis of display complexity within one experiment, and in the examination of looks to target vs. contrast as mediated by informativeness. Also, our preview analysis reveals that fixations made when the target is yet to be highlighted are not as effective as those made when it has. We hope that our findings benefit diverse areas of enquiry such as referring expression generation in NLG, referential processing, and the development of reference.

**Acknowledgements**

We are grateful to Chris Norton for preparing the eye movement data, and to Phil Vanden and Bissera Ivanova for help with experimental programming and data collection. Thanks also to Gerry Altmann and Pirita Pyykkönen-Klauck for guidance in the early stages of the project, and to two anonymous reviewers for comments on earlier versions/manuscripts. This work was funded by a British Academy Quantitative Skills grant (SQ120012) awarded to the first author.

**References**

Altmann, G. 2011. Language mediated eye movements, In S. Liversedge, I. Gilchrist & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 979-1003). Oxford: Oxford University Press.

Altmann, G. T., & Kamide, Y. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247−264.

Andersson, R., Ferreira, F., & Henderson, J. M. 2011. I see what you're saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica, 137*, 208-216.

Ariel, M. 1990. *Accessing Noun Phrase Antecedents*. London: Routledge.

Ariel, M. 2001. Accessibility theory: An overview. In Sanders, T., Schliperoord, J. and Spooren, W. (Eds.), *Text representation*. Amsterdam: John Benjamins (Human cognitive processing series), pp. 29-87.

Arnold, J.E. & Griffin, Z. 2007. The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language, 56*, 521-536.

Arnold, J.E. 2008. Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes, 23*(4), 495 -527.

Arnold, J.E., Kaiser, E., Kahn, J., & Kim, L. 2013. Information structure; Linguistic, cognitive, and processing approaches. *WIREs Cognitive Science*, Wiley. doi: 10.1002/wcs.1234

Arts, A. 2004. *Overspecification in Instructive Texts*. PhD. thesis, Tilburg University. Wolf Publishers: Nijmegen.

Arts, A., Maes, A., Noordman, L. & Jansen, C. 2011a. Overspecification facilitates object identification. *Journal of Pragmatics, 43*(1), 361-374.

Arts, A., Maes, A., Noordman, L. & Jansen, C. 2011b. Overspecification in written instruction. *Linguistics*, *49*(3), 555-574.

Audacity Team 2014. Audacity(R): Free Audio Editor and Recorder [Computer program]. Version 2.0.6 retrieved November 12th 2014 from http://audacity.sourceforge.net/

Bates, D., Maechler, M., Bolker, B., & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

Bock, J. K., Irwin, D. E., & Davidson, D. J. 2004. Putting first things first. In J. M. Henderson & F. Ferreira (Eds.), T*he interface of language, vision, and action: Eye movements and the visual world* (pp. 249-278). New York: Psychology Press.

Bock, J. K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. 2003. Minding the clock. *Journal of Memory and Language, 48,* 653-685.

Brennan, S. E. & Clark, H. H. 1996. Conceptual pacts and lexical choice in conversation*. Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*(6), 1482-1493.

Brownell, R. 2000. *Expressive one-word picture vocabulary test*. Novato, CA: Academic Therapy Publications.

Brown-Schmidt, S. & Tanenhaus, M.K. 2006. Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, *54*, 592-609.

Brown-Schmidt, S. & Tanenhaus, M.K. 2008. Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*, 643-684.

Brown-Schmidt, S., & Konopka, A.E. 2011. Experimental Approaches to Referential Domains and the On-Line Processing of Referring Expressions in Unscripted Conversation. *Information*, *2*, 302-326.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. 2008. Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*, 1122-1134.

Cave, K.R. & Bichot, N. P. 1999. Visuospatial attention: beyond a spotlight model. *Psychonomic Bulletin and Review, 6*, 204–223.

Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C.N. Li (Ed.), *Subject and topic*. New York: Academic Press Inc. pp. 25–56.

Chafe, W. 1994. *Discourse, consciousness, and time*. Chicago: Chicago University Press.

Chambers, C.G., Tanenhaus, M.K & Magnuson, J.S. 2004. Action-based affordances and syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *30*, 687-696.

Davies, C. & Arnold, J.E. (in press). Reference and informativeness: How context shapes referential choice. In C. Cummins & N. Katsos (Eds.). *Handbook of Experimental Semantics and Pragmatics.* Oxford: OUP.

Davies, C. & Katsos, N. 2010. Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua, 120*(8), 1956-1972.

Davies, C. & Kreysa, H. in prep. Look before you speak: Children's integration of visual information into informative referring expressions.

Deutsch, W. & Pechmann, T. 1982. Social-interaction and the development of definite descriptions. *Cognition, 11*(2), 159-184.

Dobel, C., Gumnior, H., Bölte, J. & Zwitserlood, P. 2007. Describing scenes hardly seen, *Acta Psychologica, 125,* 129-143, http://dx.doi.org/10.1016/j.actpsy.2006.07.004

Dunn L., D. M., Styles B., & Sewell J. 2009. *British Picture Vocabulary Scale*, 3rd Edn. (BPVS-III. . London: GL Assessment.

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*(4), 554-573.

Gordon, P. C. & Hendrick, R. 1998. The representation and processing of coreference in discourse. *Cognitive Science, 22*(4), 389-424. doi:10.1016/S0364-0213(99)80045-7

Grice, H. P. 1975. Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and semantics, 3: Speech Acts* (41-58. New York: Academic Press. Reprinted in Grice, H.P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, 22–40.

Griffin, Z. M. 2004. Why look? Reasons for eye movements related to language production. In J. Henderson & F. Ferreira, (Eds.), *The integration of language, vision, and action: Eye movements and the visual world* (pp. 213-247). New York: Taylor and Francis.

Griffin, Z. M., & Bock, K. 2000. What the eyes say about speaking. *Psychological Science, 11,* 274-279.

Grosz, B.J., Joshi, A.K. & Weinstein, S. 1995. Centering: A framework for modeling the local discourse. *Computational Linguistics, 21*(2):203–225.

Gundel, J. K., Hedberg, N., & Zacharski, R. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307.

Hanna, J. E., Tanenhaus, M. K. & Trueswell, J. C. 2003. The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*(1), 43-61.

Heller, D. & Chambers, C.G. 2014. Would a *blue kite* by any other name be just as blue? Effects of descriptive choices on subsequent referential behavior. *Journal of Memory and Language*, *70*, 53-67.

Heller, D., Grodner, D. & Tanenhaus, M.K. 2008. The role of perspective in identifying domains of reference. *Cognition, 108*, 831-836.

Hendriks, P. 2016. Cognitive modeling of individual variation in reference production and comprehension. *Frontiers in Psychology 7*:506. doi: 10.3389/fpsyg.2016.00506

Huettig, F., Rommers, J., & Meyer, A. S. 2011. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*, 151-171. doi:10.1016/j.actpsy.2010.11.003

Keysar, B., Barr, D. J., Balin, J. A. & Brauner, J. S. 2000. Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*(1), 32-38.

Koolen, R., Gatt, A., Goudbeek, M. & Krahmer, E. 2011. Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, *43*, 3231-3250.

Koolen, R., Krahmer, E., & Swerts, M. 2015. How distractor objects trigger referential overspecification: Testing the effects of visual clutter and distractor distance. *Cognitive Science,* 1-31.

Kuchinsky, S. E., Bock, J. K., & Irwin, D. E. 2011. Reversing the hands of time: Changing the mapping from seeing to saying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 748-756.

Levelt, W. J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Maes, A., Arts, A. & Noordman, L. 2004. Reference management in instructive discourse. *Discourse Processes, 37*(2), 117-144.

Mangold, R. & Pobel, R. 1988. Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology, 7,* 181-191.

Meyer, A. S. 2004. The use of eye tracking in studies of sentence generation. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 191-212). Hove: Psychology Press.

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. 2007. Early activation of object names in visual search. *Psychonomic Bulletin & Review*, *14*, 710-716

Meyer, A. S., Ouellet, M., & Häcker, C. 2008. Parallel processing of objects in a naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 982-987. doi:10.1037/0278-7393.34.4.982

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. 1998. Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2), B25-B33.

Morgan, J., & Meyer, A. S. 2005. Processing of extrafoveal objects during multiple-object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 428-442. doi:10.1037/0278-7393.31.3.428.

Mueller, S. T. 2014. PEBL: *The Psychology experiment building language* (Version 0.14) [Computer experiment programming language]. [Online]. Available: http://pebl.sourceforge.net [Accessed June 2014]

Nadig, A. S. & Sedivy, J. C. 2002. Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13*(4), 329-336.

Noveck, I. & Reboul, A. 2008. Experimental Pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences 12*(11), 425-431.

Pechmann, T. 1989. Incremental Speech Production and Referential Overspecification. *Linguistics 27*(1), 89-110.

Pickering, M. & Garrod, S. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4) 329-347. doi:10.1017/S0140525X12001495

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rabagliati, H. & Robertson, A. 2016. How do children learn to avoid referential ambiguity? Insights from eyetracking. *Journal of Memory and Language, 94*, 15-27, doi: 10.1016/j.jml.2016.09.007

Salverda, A. P., Brown, M., & Tanenhaus, M. K. 2011. A goal-based perspective on eye movements in visual world studies. *Acta Psychologica*, *137*, 172-180.

Schotter, E.R., Jia, A., Ferreira, V.S., & Rayner, K. 2014. Preview benefit in speaking occurs regardless of preview timing. *Psychonomic Bulletin & Review, 21*, 755-762.

Sedivy, J. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativeness expectations. *Journal of Psycholinguistic Research, 321*, 3-23.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G. & Carlson, G. N. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition,* 71(2), 109-147.

Seymour, H. N., Roeper, T. & De Villiers, J. G 2003. DELV-ST (*Diagnostic Evaluation of Language Variation) Screening Test.* San Antonio TX: The Psychological Corporation.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. C. 1995. Integration of visual and linguistic information in spoken language comprehension, *Science*, *258* (5217). 1632 – 1634.

Tarenskeen, S., Broersma, M., & Geurts, B. 2015. Overspecification of color, pattern, and size: Salience, absoluteness, and consistency. *Frontiers in Psychology*, 6: 1703. doi:10.3389/fpsyg.2015.01703

Trueswell, J. C., Sekerina, I., Hill, N. M. & Logrip, M. L. 1999. The kindergarten-path effect: studying online sentence processing in young children, *Cognition*, *73*(2) 89 – 134.

van der Meulen, F. F., Meyer, A. S., & Levelt, W. J. M. 2001. Eye movements during the production of nouns and pronouns. *Memory & Cognition*, *29*(3), 512-521.

van Gompel, R., Gatt, A., Krahmer, E. & van Deemter, K. 2014. Overspecification in reference: Modelling size contrast effects. Poster presented at the *20th Architectures and Mechanisms for Language Processing Conference* (AMLaP). University of Edinburgh, UK. 2014, September 6–8.

Vanlangendonck, F., Willems, R.M., Menenti, L. & Hagoort, P. 2016. An early influence of common ground during speech planning, *Language, Cognition and Neuroscience*, *31*:6, 741-750, doi: 10.1080/23273798.2016.1148747