

An early influence of common ground during speech planning

Flora Vanlangendonck, Roel M. Willems, Laura Menenti & Peter Hagoort

To cite this article: Flora Vanlangendonck, Roel M. Willems, Laura Menenti & Peter Hagoort (2016) An early influence of common ground during speech planning, *Language, Cognition and Neuroscience*, 31:6, 741-750, DOI: [10.1080/23273798.2016.1148747](https://doi.org/10.1080/23273798.2016.1148747)

To link to this article: <http://dx.doi.org/10.1080/23273798.2016.1148747>



Published online: 25 Apr 2016.



Submit your article to this journal [↗](#)



Article views: 130



View related articles [↗](#)



View Crossmark data [↗](#)

An early influence of common ground during speech planning

Flora Vanlangendonck^{a,b}, Roel M. Willems^{a,c}, Laura Menenti^c and Peter Hagoort^{a,c}

^aDonders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; ^bInternational Max Planck Research School for Language Sciences, Nijmegen, The Netherlands; ^cMax Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ABSTRACT

In order to communicate successfully, speakers have to take into account which information they share with their addressee, i.e. common ground. In the current experiment we investigated how and when common ground affects speech planning by tracking speakers' eye movements while they played a referential communication game. We found evidence that common ground exerts an early, but incomplete effect on speech planning. In addition, we did not find longer planning times when speakers had to take common ground into account, suggesting that taking common ground into account is not necessarily an effortful process. Common ground information thus appears to act as a partial constraint on language production that is integrated flexibly and efficiently in the speech planning process.

ARTICLE HISTORY

Received 29 April 2015
Accepted 21 January 2016

KEYWORDS

Language production; eye-tracking; common ground; audience design; recipient design

Introduction

A key question in language production is when and how speakers take into account which knowledge is shared between speaker and addressee (*common ground*) and which information is only available to the speaker (*privileged ground*) (Clark & Marshall, 1978). According to Grice's Maxim of Quantity, speakers should make their contributions as informative as is required for the current purpose of an interaction (Grice, 1975). Information that is not shared between interlocutors and that is uninformative should therefore not influence the language production process. However, speakers do not always successfully ignore privileged information. Imagine a situation in which a speaker sees two bottles of different sizes, the smallest of which is occluded from their addressee's point of view. If the speaker asks the addressee to hand him the mutually visible bottle, he can either describe it using a size adjective (*the large bottle*) or a bare noun (*the bottle*). According to Grice's Maxim of Quantity, the speaker should not use a size-contrasting adjective, because the smaller bottle is occluded from the listener's point of view. However, speakers in this situation regularly use a size adjective (Horton & Keysar, 1996; Wardlow Lane & Ferreira, 2008; Yoon, Koh, & Brown-Schmidt, 2012). In this case, an overspecification does not prevent the listener from identifying the intended referent. However, speakers also overspecify referring expressions when an overspecified expression can cause a misunderstanding (Wardlow Lane & Ferreira, 2008), and

when they are instructed to conceal privileged information (Wardlow Lane, Groisman, & Ferreira, 2006).

The question why speakers fail to completely ignore privileged information is still under debate, and two main theories have emerged to explain these results. According to Constraint-Based Processing models, speakers keep their addressee in mind from the earliest stages of utterance planning in a probabilistic, constraint-based way, resulting in early effects of common ground (Brennan & Hanna, 2009; Hanna, Tanenhaus, & Trueswell, 2003; Horton & Gerrig, 2002; Tanenhaus & Trueswell, 1995). In this view, privileged and common ground information act as partial constraints for language processing. When privileged information is salient, as in the example with the additional bottle above, speakers may fail to ignore it. In contrast, the Monitoring and Adjustment theory proposes that keeping track of the distinction between shared and privileged information during utterance planning is resource-intensive and often not necessary (Horton & Keysar, 1996). According to this view, speakers initially design their utterances from their own egocentric perspective, and common ground only comes into play at a later stage when speakers monitor their utterance (Keysar, Barr, & Horton, 1998). This model thus predicts that speakers initially do not distinguish between privileged and common ground information during utterance planning. Monitoring and adjusting the initial utterance plan is considered effortful and requires additional planning time (Horton & Keysar, 1996). In

CONTACT Flora Vanlangendonck ✉ flora.vanlangendonck@donders.ru.nl

Supplemental data for this article can be accessed here at <http://dx.doi.org/10.1080/23273798.2016.1148747>.

© 2016 Informa UK Limited, trading as Taylor & Francis Group

this view, speakers fail to ignore privileged information when their initial egocentric plan is not corrected.

In the current experiment, we investigated when and how shared and privileged visual information affect the production of referring expressions. Both the Constraint-Based Processing models and the Monitoring and Adjustment theory predict that privileged information *can* affect the production of referring expressions, but they make different predictions about *when* speakers take into account the distinction between shared and privileged ground. Constraint-Based Processing Models predict early effects of common ground, although privileged information may not always be successfully ignored. The Monitoring and Adjustment theory predicts effects of common ground only after an initial egocentric planning stage. Previous studies have mostly focused on speakers' adjective use. In the current experiment, we also measured eye-tracking data and planning durations in order to gain more insight into the timing of the underlying processes. We used the high temporal resolution of eye-tracking to test whether speakers distinguish between shared and privileged information during the early stages of planning a referring expression. In addition, we investigated whether ignoring privileged information during language production is resource-intensive as predicted by the Monitoring and Adjustment model by measuring speech planning durations. If speakers monitor and adapt their initial egocentric utterance plans when they design referring expressions that take into account their addressee's perspective, this should result in longer planning durations. Finally, we were interested in whether speakers avoid egocentric descriptions when those threaten communicative success. We therefore introduced two types of conditions. In the advisable audience design condition, a failure to ignore privileged information was overinformative, but the intended referent was clear. In the obligatory audience design condition, failing to ignore privileged information made it impossible for the listener to infer the intended referent. This condition, used here for the first time, thus provides a stronger test for speakers' sensitivity to communicative success than the advisable audience design condition.

Method

Participants

Twenty-two pairs of native Dutch speakers played a computerised version of a referential communication game (Keysar, Barr, Balin, & Brauner, 2000; Yoon et al., 2012). Participants did not know each other before the start of the experiment. Data from one pair were excluded

from the analyses due to experimenter error, and data from another pair were excluded because the speaker's responses could not be coded. The remaining participants consisted of 6 men and 34 women (mean age: 21.5 years, range 18–29). Half of the eye-tracking data from one pair were lost due to equipment malfunction. The remaining eye-tracking data from this pair were included in the final dataset.

Procedure

A coin toss was used to randomly assign participants to the roles of speaker and listener for the duration of the experiment. The speaker and listener were seated at separate monitors, separated by a screen to prevent them from seeing each other's monitor. Throughout the experiment, participants saw opposite sides of a 4×4 array containing objects of different sizes (Figure 1). Each array contained three closed slots on each player's side, allowing us to manipulate which objects were in common ground (i.e. visible to both participants).

Participants completed 288 trials in total. On each trial, the speaker described a specific object in the array in a way that would allow the listener to select the correct object from the array. During the first phase of the trial (3000 ms), the speaker and listener each saw their side of the array. Then the speaker was cued by means of a red circle around one of the objects, and described this object for the addressee (Figure 2). On the basis of the speaker's description, the listener selected an object by means of a mouse click. The cued object was always mutually visible. Speakers were instructed not to use descriptions referring to the position of the object in the array, such as "upper left corner" or "rightmost". Participants successfully refrained

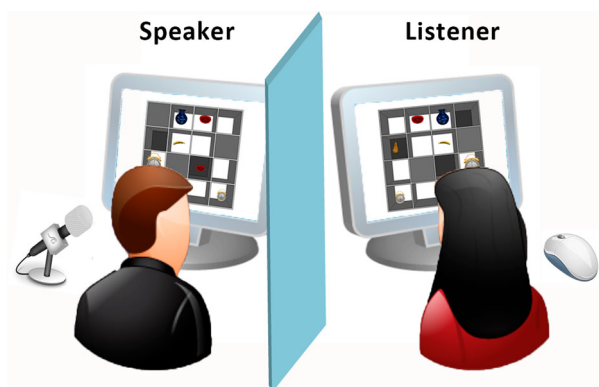


Figure 1. Overview of the set-up. Speakers and listeners viewed opposite sides of 4×4 arrays containing objects. Some objects were only visible to one of the participants, allowing us to manipulate which objects were in common ground. Speakers' responses and eye movements were recorded.

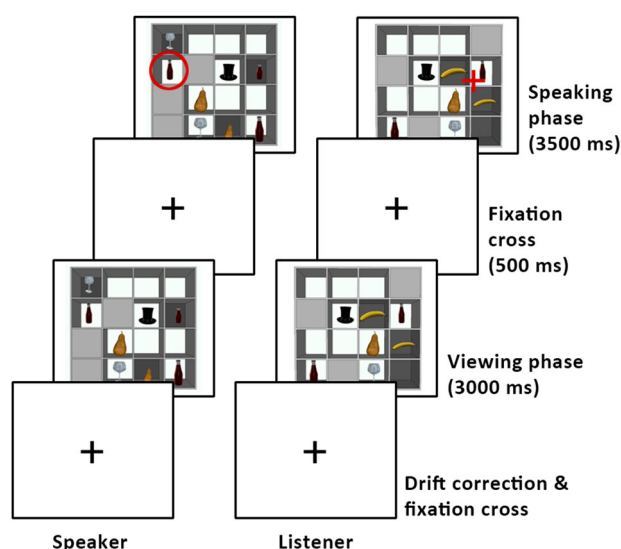


Figure 2. Trial sequence from the speaker's and the listener's point of view. In the first phase of each trial, speakers and listeners each viewed their side of the array. In the second phase of the trial, the speaker described the object that was indicated with a red circle, and the listener selected an object using the cursor.

from using such spatial descriptions (see Results section). In addition, they were instructed not to ask questions or give each other feedback during the experiment. Speakers were not given any on-screen or verbal feedback about the listener's performance during the experiment. Participants were seated next to each other in the lab, so speakers could occasionally hear the listeners' mouse clicks.

Before starting the experiment, participants practised the task together using a real array and real objects. They jointly placed the objects in the array, viewed each other's perspective, and performed the task. Then participants practised the task on the computer; during part of these test trials they were allowed to see both monitors and they could give feedback to each other.

Speakers' descriptions were recorded and their eye movements were tracked using a head-mounted EyeLink II eye-tracker (SR Research) with a 250 Hz sampling frequency. We tracked both pupil and corneal reflection whenever possible. The distance between the speaker and the monitor was approximately 57 cm. We calibrated the eye-tracker at the start of the experiment, after a break halfway through the experiment and whenever deemed necessary. Drift correction was performed before each trial.

Materials

We manipulated the number, size and visibility of the relevant objects to create six conditions (Figure 3). In the audience design conditions (left column Figure 3),

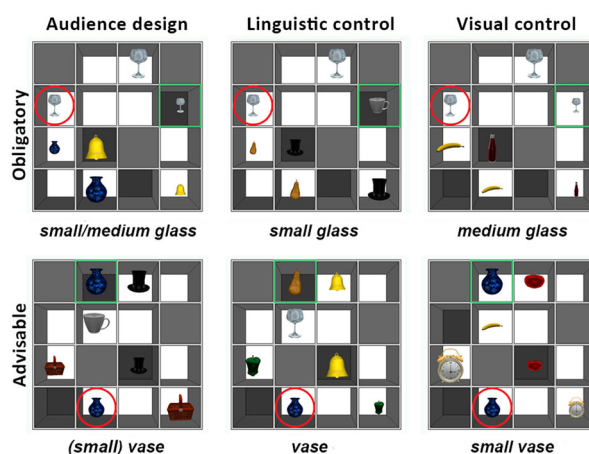


Figure 3. Overview of the six conditions from the speaker's point of view, and the expected speaker responses. In the audience design conditions, speakers can either describe the target objects (circled in red) from their own perspective ("medium glass", "small vase"), or take into account their addressee's perspective ("small glass", "vase"). There is no relevant perspective difference in the linguistic and visual control conditions. The green squares were added to the figure for clarification, and were not visible to the participants during the experiment. They indicate the objects that differ between conditions.

speakers saw an extra competitor object that the listener could not see. In the *obligatory* audience design condition, speakers saw three relevant identical objects of different sizes: one target object, one occluded competitor object and one mutually visible object. The target object was always the medium-sized object of those three objects. If speakers described this object from their own perspective, they would call it the *medium* object. In this case, their addressee had a 50% chance of selecting the wrong object. On the other hand, if speakers considered the perspective of their addressee, they would ignore the occluded object and call the medium-sized object *small* or *large*. In the *advisable* audience design condition, speakers saw two relevant identical objects of different sizes: one target object and one occluded competitor object. Given that their addressee could see the target object but not the competitor object, speakers did not have to use a contrasting size adjective. However, unlike in the obligatory audience design condition, listeners were still able to select the correct object if the speaker did not consider the perspective difference.

We created two types of control conditions in which speakers and listeners saw the same number of relevant objects. In the linguistic control conditions (middle column Figure 3), the occluded object was replaced by another, unrelated object. As a result, speakers saw one relevant object fewer in these conditions than in the audience design conditions. We called these *linguistic* control conditions, because speakers were expected to

produce the same description on these trials as on successful trials in the audience design conditions. In the visual control conditions (right column Figure 3), the object that was occluded in the audience design conditions was visible to both participants. As a result, speakers and listeners could both see all relevant objects. We called these *visual* control conditions, because speakers see the same number of relevant objects as in the audience design conditions. Neither of these control conditions required the speaker to take into account the perspective difference with their addressee in order to communicate successfully.

We created 12 different empty virtual arrays. The arrays were filled with 6–8 objects chosen from a total of 22 objects. Objects were selected from the Object Databank, stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University (<http://www.tarrlab.org/>). Each object could appear in four different sizes to make sure that participants could not rely on absolute size. Depending on the condition, speakers saw one, two or three relevant objects of the same type but of different sizes. The remaining objects were fillers that also appeared in sets of one, two or three objects of the same type to make sure participants could not predict which objects would be relevant. We made sure that the speaker and the listener always saw the same total number of objects in a trial by adding additional filler objects to the occluded slots if needed. We have created a visual overview depicting the trial construction, which can be found in the supplementary materials. For each speaker, we created 288 trials by constructing 48 triplets of obligatory trials and 48 triplets of advisable trials. Each triplet consisted of one audience design trial, one linguistic control trial and one visual control trial. Figure 3 shows examples of one obligatory triplet (top row) and one advisable triplet (bottom row). The target and competitor objects per triplet were identical in type, size and location, with the exception of the occluded object in the linguistic control condition. The number, proportion and location of filler objects were kept constant within each triplet, but we varied the type and the size of the filler objects in order to avoid memory effects. As can be seen in Figure 3, the audience design and the linguistic control trials in a triplet were created using the same array. The array used in the visual control condition differed from this array in the location of one closed slot. Each of the 12 virtual arrays was used to create 4 obligatory and 4 advisable triplets of trials. The only thing these triplets had in common was the array that was used as a starting point. Target object–location pairings were only repeated within a triplet, i.e. three times throughout the experiment.

During the experiment, trials were presented in blocks of six trials that were created using the same array. Each block contained no more than one trial from a triplet. The order of the trials within each block was randomised, and we randomised the blocks of trials so that neighbouring blocks did not use the same array. We created unique stimulus lists for each participant, detailing which object should appear where in each array. Based on these lists, the stimulus pictures, as presented to the participants, were assembled online in Presentation software.

Data analysis

We coded the sound files offline for adjective use. For the obligatory trials, we coded the use of *small/large* and *medium* adjectives (*klein/groot* and *middelgroot* in Dutch). For the advisable trials, we coded the use of bare nouns and *small/large* responses. In addition, we coded errors, false starts, repairs, speech unrelated to the task, and responses exceeding the response interval of 3500 ms, and removed trials that contained any of these from further analysis (10.45% of trials). Listener performance was calculated by determining whether listeners clicked in the right slot within the 3500 ms-response interval. Finally, we computed planning durations (from cue onset until speech onset) and speaking durations (from speech onset until speech offset) from the sound files using a custom-made Matlab script. The script filtered and smoothed the waveform, determined an initial speech onset and offset by looking for sustained periods of silence and speech in the spectrogram and then calculated a more precise speech onset and offset by using increasingly smaller windows and lower thresholds. Informal inspection of the speech waveforms revealed that these calculations were accurate (example waveforms are included in the supplementary materials).

The statistical analyses were performed in R version 3.0.3 (R Core Team, 2014). We used the `lmer` function from the `lme4` package version_0.999999-4 (Bates, Mächler, Bolker, & Walker, 2015) to fit the linear mixed models and we built up the models from a simple to a more complex model. The linear mixed models (Baayen, Davidson, & Bates, 2008) included a random intercept for subjects, a by-subjects random slope for condition, a random intercept for target object indicating which of the 22 possible objects was used as target object, and condition and number of objects as fixed effects. Models were estimated using maximum likelihood estimation. The factor *condition* had three levels (audience design, linguistic control and visual control) and the factor *number of objects* had three levels (six, seven or eight objects). We used the same model in

every analysis, except that the reference distribution was changed depending on the dependent measure: Poisson for count data (number of fixations per trial), binomial for binary data (adjective use) and Gaussian for continuous data (planning durations, speaking durations and fixation durations). Obligatory and advisable data were analysed separately. The obligatory conditions contained more relevant objects than the advisable conditions, so we could not compare them directly.

The main effect of condition was tested by comparing a complete model containing all random and fixed effects to a model in which the fixed effect of condition was removed (the by-subjects random slope for condition remained in the model). For each dependent measure, two planned contrasts were performed in order to compare the audience design condition with each of the control conditions. We used the `glht` function from the `multcomp` package version 1.3-2 (Hothorn, Bretz, & Westfall, 2008) to run these planned comparisons. More information about the data analysis can be found in the supplementary materials.

Results

Adjective use

We computed the percentage of each type of response for the different conditions to find out how often speakers took their addressee's perspective into account when producing a referring expression. For the obligatory trials, we treated the presence of the adjectives *small* or *large* as opposed to *medium* as a binary dependent variable; for the advisable conditions, we treated the presence of a bare noun as opposed to a size adjective (*small* or *large*) in the speakers' descriptions as a binary dependent variable.

Speakers adapted their language use to their addressee's perspective in the majority of audience design trials (Figure 4), yet they failed to ignore privileged information on all trials. In the obligatory audience design condition, speakers mainly produced *small/large* responses (89.88%), although they also used *medium* responses (10.12%). *Medium* responses are accurate descriptions of the target object from the speakers' perspective, but do not allow the listener to uniquely identify the intended referent. As expected, speakers mostly produced utterances that contained a *small/large* size adjective (97.32%) in the linguistic control condition and mainly *medium* responses (98.59%) in the visual control condition. We found a significant main effect of condition on the use of *small/large* size adjectives ($\chi^2(2) = 41.71, p < .001$). Both the difference in adjective use between the obligatory audience design condition and the linguistic

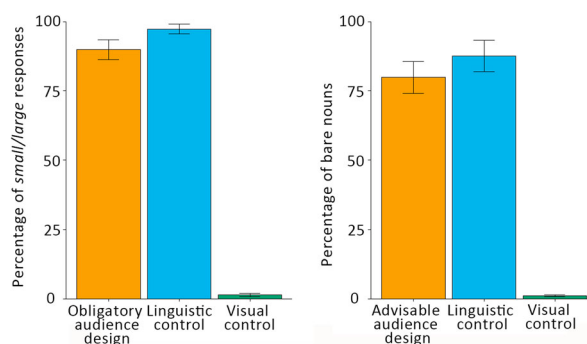


Figure 4. The percentages of speakers' small/large and bare noun responses for the obligatory (left) and advisable data (right). Bars indicate standard error of the mean. Speakers were expected to produce such responses in the audience design conditions if they took their addressee's perspective into account.

control condition ($b = 3.75, SE = 0.51, p < .001$) and the difference in adjective use between the obligatory audience design condition and the visual control condition were significant ($b = -8.77, SE = 0.72, p < .001$).

In the advisable audience design condition, speakers generally took their addressee's perspective into account. They mostly produced bare nouns (79.84%), although they also produced *small/large* responses (20.16%). In the linguistic control condition, speakers mainly produced bare nouns (87.62%), and in the visual control condition they predominantly used *small/large* responses (98.90%). We found a significant main effect of condition on the use of bare nouns ($\chi^2(2) = 39.88, p < .001$). Both the difference between the advisable audience design condition and the linguistic control condition ($b = 1.64, SE = 0.28, p < .001$) and the difference between the advisable audience design condition and the visual control condition were significant ($b = -6.60, SE = 0.60, p < .001$).

Overall listener accuracy was high: listeners selected the intended object on 93.60% of the obligatory trials and 97.70% on the advisable trials. As a manipulation check we looked specifically at listeners' performance on audience design trials on which speakers described the target object without taking their addressee's perspective into account (i.e. saying *medium* in the obligatory audience design trials and using a size adjective in the advisable audience design trials). This was a small subset of trials, unequally divided over participant pairs (obligatory data: 2.98% of trials; advisable data: 6.14% of trials). On obligatory audience design trials, listeners only selected the correct object in 36% of these ambiguous trials; on advisable audience design trials, listeners were successful in 98.15% of trials. This confirms that our advisable/obligatory manipulation worked as expected: not taking the addressee's perspective into

account forced the listener to guess the referent in the obligatory audience design trials, but not in the advisable audience design trials.

Planning and speaking durations

We computed speakers' planning durations (from cue onset to speech onset) based on the sound recordings to test whether ignoring privileged information during utterance planning is resource-intensive. We found no significant main effect of condition in both the obligatory data ($\chi^2(2) = 1.82, p = .40$) and the advisable data ($\chi^2(2) = 0.84, p = .66$). Figures of the mean planning durations per condition are included in the supplementary materials.

We also calculated speaking durations (from speech onset to speech offset). If ignoring privileged information is a resource-intensive process, this might also lead to longer speaking durations. We found a significant main effect of condition on speaking durations, both for the obligatory ($\chi^2(2) = 26.23, p < .001$) and the advisable data ($\chi^2(2) = 45.65, p < .001$). No significant difference in speaking time was found between the obligatory audience design condition and the linguistic control condition ($b = -17.76, SE = 8.39, p = .067$). Speaking durations were shorter in the obligatory audience design condition than in the visual control condition ($b = 86.54, SE = 12.38, p < .001$), most likely because the Dutch word *middelgroot* has more syllables than the other size adjectives (*klein* and *groot*). We found significant differences in speaking duration between the advisable audience design and the linguistic control condition ($b = -26.61, SE = 8.96, p < .01$) and between the advisable audience design condition and the visual control condition ($b = 224.43, SE = 17.52, p < .001$). However, when we removed the advisable audience design trials in which the speaker did not take the addressee's perspective into account (i.e. the speaker produced a size adjective), the difference between the advisable audience design condition and the linguistic control condition disappeared ($b = 15.95, SE = 14.34, p = .46$). Figures of the mean speaking durations per condition are included in the supplementary materials.

Number of fixations per trial

In order to measure when privileged information affects the production of referring expressions, we computed the mean number of fixations on the occluded competitor object in the audience design conditions and on the objects in the same locations in the associated control conditions (the objects in the green squares in Figure 3). We focused our analysis specifically on the planning duration window, i.e. in each trial, the time

between the moment the array with the cued object was presented and the time the speaker started talking. For both the obligatory and the advisable data, we found that speakers fixated an occluded competitor object (audience design conditions) more than an unrelated control object (linguistic control), but less than a mutually visible competitor object (visual control). Figures 5 and 6 show how the proportion of fixations on the target and occluded object changes over time and give an overview of the number of fixations per trial. Figures 5(a) and 6(a) clearly show that speakers initially fixate the target object they have to describe, and then consider the other objects in the array.

We found a significant main effect of condition on the number of fixations per trial in the obligatory data, ($\chi^2(2) = 52.28, p < .001$). We found significantly more fixations per trial on the occluded competitor object in the obligatory audience design condition compared to the occluded control object in the linguistic control condition ($b = -1.57, SE = 0.15, p < .001$). We found significantly less fixations per trial on the occluded competitor object in the obligatory audience design condition compared to the mutually visible object in the visual control condition ($b = 0.89, SE = 0.12, p < .001$). To rule out that these differences were driven by the small percentage of obligatory control trials on which speakers responded with a *medium* response, we ran the same analysis including only trials in which speakers' responses took into account the listener's perspective (i.e. a *medium* response in the visual control condition, a *small/large* response in the audience design and linguistic control conditions). The differences remained significant in this analysis. Across the obligatory conditions, the mean number of fixations per trial is relatively low, even when the competitor object was relevant (visual control condition). Some participants reported that they paid special attention to objects that appeared in triplets during the first half of the trials, which may have allowed them to plan their utterance after the cue without additional fixations on the other objects.

We observed a similar pattern in the advisable data. We found a significant main effect of condition on the number of fixations per trial, ($\chi^2(2) = 41.99, p < .001$). We found a significant difference in the mean number of fixations per trial between the advisable audience design condition and the linguistic control condition ($b = -1.44, SE = 0.11, p < .001$) and between the advisable audience design condition and the visual control condition ($b = 0.31, SE = 0.063, p < .001$). This pattern remained even when we only included the trials in which speakers' responses took into account the listener's perspective (i.e. an adjective in the visual control condition, a bare noun in the advisable audience

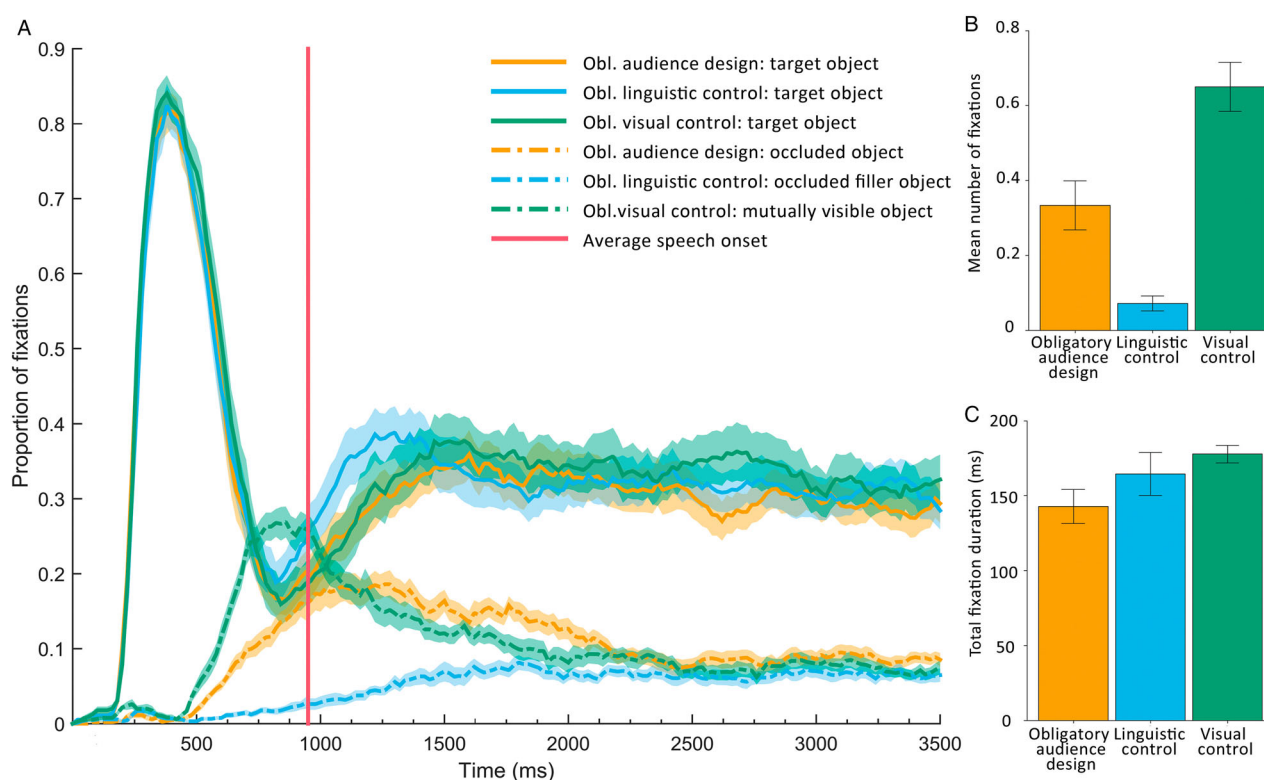


Figure 5. Eye-tracking results for the obligatory data. (A) Proportions of speakers' fixations to the target and occluded objects from the moment speakers were cued until the end of the trial. Solid lines indicate fixations to the target object; dashed lines represent fixations to the occluded competitor object, the occluded competitor object or the mutually visible competitor object. Bounded lines indicate the standard errors. (B) Mean number of fixations (C) and total fixation time during the planning duration window on the occluded competitor object in the obligatory audience design condition, on the occluded filler object in the linguistic control condition and on the mutually visible competitor object in the visual control condition.

design and linguistic control conditions). These results show that speakers already distinguish between common and privileged ground while planning their utterance, although they do not completely ignore privileged information.

Total fixation duration per trial

We computed the total duration of fixations on the occluded competitor object in the audience design conditions and on the objects in the same locations in the associated control conditions. We again limited the analysis to fixations during the planning duration window. Fixations that exceeded the planning duration window were truncated and we treated trials without fixations on the occluded object as missing data. Speakers fixated the unrelated occluded object in the linguistic control condition in only 6.49% of obligatory linguistic control condition trials, and on 16.08% of advisable linguistic control trials. Given that there were so few of these trials, we did not contrast the audience design conditions and the linguistic control conditions for this dependent variable. An overview

of the average total fixation durations can be found in Figures 5(c) and 6(c).

Speakers looked longer at mutually visible objects (visual control conditions) than at objects that were occluded from the listener's point of view (audience design conditions). We found significant main effects of condition in both the obligatory ($\chi^2(2) = 14.15, p < .001$) and the advisable data ($\chi^2(2) = 27.12, p < .001$). We found a significant difference in total fixation duration between the obligatory audience design condition and the visual control condition ($b = 31.82, SE = 9.41, p < .01$). Similarly, speakers spent less time looking at the occluded competitor object in the advisable audience design condition than at the mutually visible object in the visual control condition ($b = 22.53, SE = 7.16, p < .01$). These results again support the finding that speakers already treat common and privileged information differently while planning their utterance.

Discussion

In this experiment, we investigated the effect of privileged information on the production of referring

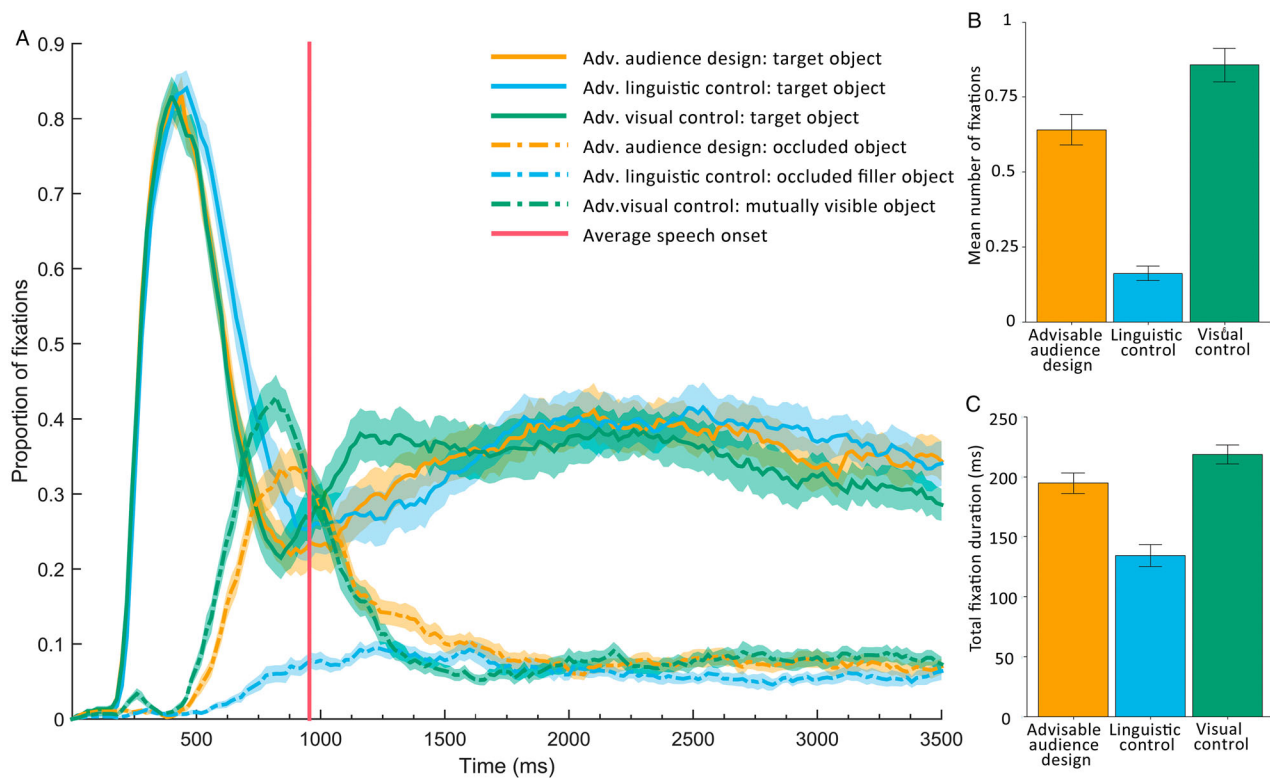


Figure 6. Eye-tracking results for the advisable data. (A) Proportions of speakers' fixations to the target and occluded objects from the moment speakers were cued until the end of the trial. Solid lines indicate fixations to the target object; dashed lines represent fixations to the occluded competitor object, the occluded competitor object or the mutually visible competitor object. Bounded lines indicate the standard errors. (B) Mean number of fixations (C) and total fixation time during the planning duration window on the occluded competitor object in the advisable audience design condition, on the occluded filler object in the linguistic control condition and on the mutually visible competitor object in the visual control condition.

expressions. Speakers mainly produced referring expressions that took into account their addressee's visual perspective. However, the availability of privileged information also led to the production of utterances that did not take this perspective difference into account. These findings are in line with previous studies using similar paradigms (Wardlow Lane et al., 2006; Yoon et al., 2012). Interestingly, speakers even failed to completely ignore privileged information when it harmed communication (i.e. the obligatory audience design condition). In a previous study, Wardlow Lane and Ferreira (2008) also found that speakers were not able to fully ignore privileged information even when it threatened communicative success. However, in their study target objects were in privileged ground in certain filler trials, meaning that speakers could not systematically ignore privileged ground information. In the current experiment, speakers could fully ignore occluded objects throughout the experiment, yet even threats to communicative success did not prevent occasional interference of privileged information.

In addition to studying the form of speakers' referring expressions, we collected planning durations and eye-

tracking data to address when and how privileged information affects the production of referring expressions. Speakers fixated occluded competitor objects less than relevant competitor objects both in terms of the number of fixations and the total fixation duration. This suggests that speakers did not initially treat privileged information in an egocentric manner as predicted by the Monitoring and Adjustment hypothesis, but that they take into account which objects are visible to their communication partner. More specifically, speakers used the information that was available to them in the first half of each trial to distinguish between common and privileged ground when they were planning their referring expressions. Despite this early distinction, speakers did not fully ignore privileged information when planning their referring expression. We found more fixations on the occluded target objects than on unrelated occluded objects. This shows that speakers' general success at producing utterances that took into account their addressee's perspective was not the result of fully ignoring all occluded objects. Fixations to the occluded competitor object might result from speakers' occasional failure to ignore their own perspective.

Alternatively, speakers may deliberately fixate the occluded object to check whether it is in common ground, for example, because they failed to look at the occluded object during the viewing phase or as part of monitoring during speech production. Based on our data, we cannot exclude either option. To test the prediction of the Monitoring and Adjustment theory that audience design should lead to additional planning time (Horton & Keysar, 1996), we computed and compared planning durations. However, unlike Horton and Keysar (1996), we did not find any significant differences in planning duration between the audience design conditions and the control conditions. One important difference between these experiments is that in our experiment participants had the opportunity to view the array and objects before they had to describe the target object. We included the viewing phase, because in most real-life communicative settings, speakers also have the opportunity to process their environment before they start talking. The viewing phase may have facilitated the speech planning process by allowing speakers to compute which objects were in common ground before they started planning their utterances. In addition, we repeated the same 22 images scaled to different sizes during the experiment, which may have allowed speakers to learn to identify absolute object size and may have thus facilitated target object naming over the course of the experiment. However, we consider this possibility unlikely given that the objects had different sizes to begin with and could appear on screen in four possible sizes, making it hard to determine absolute object size. Taken together, our results show that common ground does not function as a complete constraint on the production of referring expressions, but does exert an early effect during utterance planning. Speakers tried to use the information that was available to them during the first half of each trial to restrict common ground when they were planning their utterance, as reflected in the lower number of fixations on the occluded object during planning in the second half of the trials. When speakers fixated the occluded object during planning, this often did not prevent them from tailoring their referring expression for their addressee. These results support Constraint-Based Processing models. In this view, common and privileged ground act as probabilistic constraints to guide language processing in combination with other constraints such as context (Brown-Schmidt & Hanna, 2011). Given the lack of differences in planning durations and speakers' relative success at tailoring their referring expressions for their addressee, weighing the available shared and privileged information appears to be a relatively efficient process. The finding that common ground functions

merely as a partial constraint is perhaps best understandable when you consider that in many communicative situations the goal of the interaction is exactly to provide privileged information (e.g. giving directions or responding to questions). Maybe speakers' attention to privileged objects should therefore be thought of as a useful feature of communicative language processing rather than a failure to ignore irrelevant information. In addition, referring expressions that contain uninformative privileged information often do not prevent the addressee from identifying the intended referent, although they may temporarily confuse the addressee (Engelhardt, Bailey, & Ferreira, 2006).

Our findings of early and partial effects of common ground on language *production* are similar to previous findings in language *comprehension* research showing early effects of common ground in reference resolution (Hanna et al., 2003; Heller, Grodner, & Tanenhaus, 2008; Nadig & Sedivy, 2002). This suggests that common ground may affect language production and language comprehension in similar ways.

Conclusion

In line with previous studies using similar paradigms (Wardlow Lane et al., 2006; Yoon et al., 2012), we found that speakers cannot completely ignore privileged information during language production and that the availability of privileged information can lead to the production of utterances that do not take the addressee's visual perspective into account. However, we found no evidence that adapting to your addressee's visual perspective requires additional planning time, suggesting that audience design is not necessarily effortful. Combined, these results suggest that the available privileged and shared information can efficiently constrain language production. These findings support Constraint-Based Processing models that predict that common and privileged information is incorporated into language processing in a probabilistic fashion. Our results show that common ground does not exert an all-or-nothing influence on language production. Instead, common ground influences the production process as a partial constraint already during utterance planning.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and

- items. *Journal of Memory and Language*, 59(4), 390–412. doi:10.1016/j.jml.2007.12.005
- Bates, D., Mächler, M., Bolker, B.M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01
- Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science*, 1(2), 274–291. doi:10.1111/j.1756-8765.2009.01019.x
- Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialog and Discourse*, 2(1), 11–33. doi:10.5087/dad.2011.102
- Clark, H. H., & Marshall, C. R. (1978). Reference diaries. In D. L. Waltz (Ed.), *Theoretical issues in natural language processing* (Vol. 2, pp. 57–63). New York, NY: Association for Computing Machinery.
- 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. doi:10.1016/j.jml.2005.12.009
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech arts* (pp. 41–58). New York, NY: Academic Press.
- 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. doi:10.1016/S0749-596X(03)00022-6
- Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, 108(3), 831–836. doi:10.1016/j.cognition.2008.04.008
- Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, 47(4), 589–606. doi:10.1016/S0749-596X(02)00019-0
- Horton, W.S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59(1), 91–117. doi:10.1016/0010-0277(96)81418-1
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363. doi:10.1002/bimj.200810425
- 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. doi:10.1111/1467-9280.00211
- Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, 7(2), 46–50. doi:10.1111/1467-8721.ep13175613
- 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. doi:10.1111/j.0956-7976.2002.00460.x
- R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL <http://www.R-project.org/>.
- Tanenhaus, M., & Trueswell, J. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of cognition and perception: Vol. 11. Speech and language* (pp. 217–262). San Diego, CA: Academic Press.
- Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1466–1481. doi:10.1037/a0013353
- Wardlow Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological Science*, 17(4), 273–277. doi:10.1111/j.1467-9280.2006.01697.x
- Yoon, S. O., Koh, S., & Brown-Schmidt. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, 19(4), 699–707. doi:10.3758/s13423-012-0262-6