COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2014) 1–23

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12205

How Cognitive Load Influences Speakers' Choice of Referring Expressions

Jorrig Vogels,[a] Emiel Krahmer,[b] Alfons Maes[b]

[a]*Department of Computational Linguistics and Phonetics and Department of Psychology, Saarland University, Saarbrücken*
[b]*Tilburg Center for Cognition and Communication (TiCC), Tilburg University*

Received 10 July 2013; received in revised form 18 July 2014; accepted 18 July 2014

Abstract

We report on two experiments investigating the effect of an increased cognitive load for speakers on the choice of referring expressions. Speakers produced story continuations to addressees, in which they referred to characters that were either salient or non-salient in the discourse. In Experiment 1, referents that were salient for the speaker were non-salient for the addressee, and vice versa. In Experiment 2, all discourse information was shared between speaker and addressee. Cognitive load was manipulated by the presence or absence of a secondary task for the speaker. The results show that speakers under load are more likely to produce pronouns, at least when referring to less salient referents. We take this finding as evidence that speakers under load have more difficulties taking discourse salience into account, resulting in the use of expressions that are more economical for themselves.

Keywords: Cognitive load; Reference; Pronouns; Language production; Accessibility; Perspective taking

1. Introduction

When speakers refer to something that has been mentioned before, they can choose between different types of referring expressions, such as a definite description (e.g., *the girl*) or a pronoun (e.g., *she*). Traditionally, the speaker's choice of a referring expression in discourse has been assumed to be tailored for the addressee (e.g., Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993). According to this view, speakers make assumptions about the cognitive status of the referent in the mind of their addressee, for example, whether

Correspondence should be sent to Jorrig Vogels, Department of Computational Linguistics and Phonetics and Department of Psychology, P.O. Box 15 11 50, 66041 Saarbrücken, Germany. E-mail: jorrig@coli.uni-saarland.de

the referent's mental representation is highly accessible for the addressee or not. The most important factor in determining this cognitive status is assumed to be the salience of the referent in the preceding discourse. For example, if the referent was the topic of the previous sentence, it can be assumed to be salient in the discourse (e.g., Givón, 1983; Grosz, Joshi, & Weinstein, 1995). As a result, its mental representation is likely to be highly accessible in the addressee's discourse model. That is, it can be easily retrieved from memory. Therefore, the referent does not need an elaborate description to be retrieved by the addressee, and the speaker can suffice with an attenuated expression, such as a pronoun. Because, so the classical reasoning goes (e.g., Gundel et al., 1993), the addressee knows that the speaker would have used a more elaborate expression if she had a less salient referent in mind, the use of a pronoun aids the addressee's interpretation. This is in line with the idea that cooperative speakers obey Grice's Maxim of Quantity (Grice, 1975): Speakers choose referring expressions that are as informative as required for the addressee to pick out the correct referent, but not more informative than required.

More recently, it has been suggested that the choice of a referring expression may also be influenced by speaker-internal constraints (e.g., Arnold, 2008; Arnold, Bennetto, & Diehl, 2009; Arnold & Griffin, 2007). Speakers are not always monitoring the communicative needs of their addressees (e.g., Dell & Brown, 1991; Engelhardt, Bailey, & Ferreira, 2006; Wardlow Lane, Groisman, & Ferreira, 2006). One reason is that the language production system is constrained by the speaker's attention resources and working memory capacity. Since these resources are limited (Baddeley, 1986), speakers do not have unrestricted processing capacity to keep track of all elements in the discourse and to calculate the accessibility of referents for the addressee. In addition, it has become clear that even when there is sufficient processing capacity, people do not always initially take the perspective of their conversation partners into account in producing referring expressions (e.g., Bard et al., 2000; Fukumura & Van Gompel, 2012; Gann & Barr, 2014), although they might do this eventually, in a later stage of processing (e.g., Dell & Brown, 1991; Horton & Keysar, 1996; Keysar, Barr, & Horton, 1998).

It is less clear, however, how exactly speaker-internal constraints affect reference production. In this study, we experimentally investigate how an increased memory load on the part of the speaker influences how speakers choose between attenuated expressions such as pronouns and more elaborate expressions such as full noun phrases. Assuming that such cognitive load taps into the language production process, manipulating it gives more insight in the mechanisms underlying referential choices. The next sections discuss different hypotheses with respect to the possible effects of cognitive load on the choice of referring expression. On the one hand, if speakers are choosing referring expressions based on their assumptions about the referent's accessibility in the addressee's model of the discourse, an increased cognitive load may hinder the ability of speakers to make these assumptions (Section 2). On the other hand, if speakers are choosing referring expressions based on how accessible the referent is for themselves, increased cognitive load may affect the referent's representation in speakers' own discourse models (Section 3).

2. Hypothesis 1: Cognitive load makes reference more egocentric

If choosing referring expressions involves taking into account how accessible the referent is for the addressee, an increased cognitive load may make this harder. Different models of audience design have been proposed to account for the fact that speakers are not always monitoring the knowledge of their addressee. According to the Monitoring and Adjustment model (Horton & Keysar, 1996), for example, speakers initially plan their utterances egocentrically, that is, without taking into account common ground with their addressees. A subsequent process then checks this initial plan for errors, such as whether it is relying on information that is not accessible for the addressee, and adjusts it when necessary. Since this monitoring involves an additional step in processing, it is predicted to take up more time and memory resources. Indeed, Horton and Keysar (1996) found that speakers took into account the addressee's perspective when they had to choose whether or not to include an adjective in their referring expressions. However, they were less able to do this when they were under time pressure. In that case, speakers more often based their utterances only on information accessible to them.

Another model of audience design is the Dual Process model (Bard & Aylett, 2005; Bard et al., 2000). This model makes a distinction between automatic processes that only take into account the speaker's knowledge, and more effortful processes that build inferences about the addressee's knowledge. These inferential processes compete for attention with task demands: The more attention a task requires, the less speakers will take the addressee's knowledge into account.

Thus, both the Monitoring and Adjustment model and the Dual Process model suggest that restrictions on the processing capacity needed for audience design may make references more egocentric. However, it is not completely clear what it means for speakers to be egocentric when they choose between pronouns and full noun phrases in discourse. On the one hand, it could mean that speakers base the choice of referring expressions on their own model of the discourse rather than on assumptions about the addressee's discourse model (e.g., Bard & Aylett, 2005; Fukumura & Van Gompel, 2012). On the other hand, referring egocentrically could mean that speakers are inclined to use those referring expressions that are easiest for them to produce (e.g., pronouns; Hendriks, Englert, Wubs, & Hoeks, 2008).

If the choice of referring expression becomes more based on the speaker's own model of the discourse when memory load is high, speakers are expected to be more likely to use a pronoun when the referent is highly accessible for them, and a full noun phrase when the referent is less accessible for them, rather than consider accessibility from the addressee's perspective. In many cases, of course, this will not cause problems, since speakers and addressees tend to have closely aligned discourse models (e.g., Arnold, 2008; Pickering & Garrod, 2004); a referent that is highly accessible in the speaker's discourse model is typically also highly accessible in the addressee's discourse model. However, when the speaker's and the addressee's perspectives differ—for instance, because the addressee did not hear part of the preceding discourse—speakers under load might be

inclined to use pronouns if the referent is salient in their own discourse model but not necessarily salient in the addressee's discourse model. Conversely, they might be inclined to use full noun phrases if the referent is not salient in their own discourse model but salient in the addressee's discourse model.

To investigate whether speakers are taking into account their addressee's perspective when choosing referring expressions, Fukumura and Van Gompel (2012) conducted a story completion experiment in which the sentence directly preceding the speaker's continuation was in privileged ground, that is, it was only heard by the speaker over headphones. This privileged sentence either made the referent to be described in the continuation (the target referent) discourse salient, or it made a competitor referent salient. In both cases, the target referent was not salient for the addressee. Therefore, if speakers were taking into account the addressee's discourse model, they should use a full noun phrase to refer to the target referent, irrespective of the content of the privileged context sentence. If speakers were using their own discourse model, they should use more pronouns when the target referent was made salient in the privileged sentence than when it was not. The results of this study showed the latter pattern, suggesting that speakers were more likely to follow their own discourse model than to take into account their addressee's perspective. Still, speakers were more likely to use full noun phrases than in another condition in which all discourse context was in common ground. This suggests that at least some audience design was going on. One possibility is that under load, speakers are more likely to abandon such audience design and choose referring expressions based on their own discourse model.

Referring egocentrically may also mean that speakers choose more economical expressions overall. Due to their reduced phonological and semantic content, pronouns may be more economical in terms of processing costs than full noun phrases (Almor, 1999; Burzio, 1998; Levinson, 1987). Therefore, speakers may inherently prefer to produce pronouns over more specific expressions. On this view, given that speakers themselves know what they are referring to, producing more elaborate expressions is simply not beneficial for speakers. Any expression that is less economical than a pronoun may thus be considered as somehow tailored for a (potentially hypothetical) addressee (Hendriks, Koster, & Hoeks, 2014). If this addressee-oriented process is cognitively effortful, the preference to use pronouns may be reinforced when speakers do not have enough processing capacity to take into account the knowledge of the addressee. Indeed, studies have found that speakers with a low working memory capacity (children, elderly) are more likely to use pronouns in contexts in which the referent is not salient for the addressee (and hence a more specific expression would normally have been appropriate; Hendriks et al., 2008, 2014; Wubs, Hendriks, Hoeks, & Koster, 2009; see also Almor, Kempler, MacDonald, Andersen, & Tyler, 1999).

In sum, our first hypothesis is that that cognitive load will make speakers more egocentric. This could result in either choosing referring expressions based on the speaker's own discourse model (i.e., producing pronouns for referents salient for the speaker and full noun phrases for non-salient referents) or generally using more economical expressions (i.e., pronouns).

3. Hypothesis 2: Cognitive load affects the speaker's own discourse model

If speakers are using their own discourse model when choosing referring expressions, rather than taking into account the discourse model of the addressee, regardless of whether they have enough processing capacity for that, cognitive load may directly affect the accessibility of mental representations in the speaker's own memory (Arnold, 2010). Indications that the activation of discourse elements in the speaker's own discourse model influences the choice of referring expression come from studies that manipulate the speaker's attention resources. For example, Arnold and Griffin (2007; see also Fukumura, Van Gompel, & Pickering, 2010) conducted a story completion experiment in which they varied the number of possible referents in the discourse. They found that speakers used fewer pronouns when a referential competitor was present, even though pronouns were never ambiguous and the target referent was salient in the discourse. Hence, a pronoun reference could have been easily resolved by the addressee. Speakers have also been found to use fewer attenuated expressions when they are distracted by another task (Rosa & Arnold, 2011), and when they are either disfluent or planning longer utterances, which are both considered indications for an increased cognitive load (Arnold et al., 2009). These findings have been explained as evidence for a decrease in the accessibility of the referent in the speaker's own discourse model when attentional resources have to be spread over multiple possible referents or multiple (effortful) tasks. Thus, on this view, speakers with decreased cognitive resources are less likely to use attenuated expressions, because the activation of the referents in their own discourse model is reduced.

In sum, our second hypothesis is that speakers base their choice of referring expressions on the salience of the referent in their own model of the discourse. In this case, an increased cognitive load may reduce referent accessibility, resulting in more elaborate expressions.

4. Predictions and experimental design

We have formulated two, not necessarily mutually exclusive, hypotheses with respect to the effect of an increased cognitive load on speakers' choice of referring expressions. First, when they are under load, speakers may be less likely to take into account the addressee's perspective, causing them to either fall back on their own discourse model or use more economical expressions. Second, cognitive load may affect mental representations in the speaker's own discourse model, resulting in a reduced accessibility of the referents, and hence in more elaborate referring expressions.

To tease these possible effects of cognitive load apart, the speaker's and the addressee's perspectives with respect to the discourse salience of the referent should be dissociated. In this way, it can be determined whether speakers under load have more difficulties to choose referring expressions based on their addressee's perspective. In addition, references to both salient and non-salient referents should be investigated, because

cognitive load might affect these differently. Indeed, most studies that found a decrease in pronoun use with increased cognitive load (e.g., Arnold & Griffin, 2007; Fukumura et al., 2010; Rosa & Arnold, 2011) investigated only contexts in which the target referent was the subject of the preceding sentence and was therefore salient for the speaker (and the addressee, whether hypothetical or not). In contrast, studies that found an increase in pronoun use in speakers with a lower working memory capacity either investigated only non-salient referents (i.e., after a topic shift; Hendriks et al., 2008; Wubs et al., 2009) or did not control for discourse salience (Almor et al., 1999). In addition, none of these studies dissociated the speaker's and addressee's perspectives, by which it remains unclear whether cognitive load was affecting referent accessibility in the speaker's own discourse model or the speaker's assumptions about referent accessibility in the addressee's discourse model. This study was conducted to shed more light on this question.

We conducted two story completion experiments in Dutch, in which we manipulated the cognitive load of the speaker, as well as referent salience. In the experiments, speakers were presented with pairs of pictures showing two characters. Following two context sentences, speakers produced a continuation for an addressee, in which they referred to one of these characters, which was either salient or non-salient in the discourse. We manipulated cognitive load by having speakers conduct a verbal memory task while telling the stories in one half of the experiment. We used a verbal rather than a visual secondary task (cf. Rosa & Arnold, 2011) to make sure that it would interfere with memorizing or attending to discourse information rather than with visually attending to the characters in the pictures (e.g., Kellogg, Olive, & Piolat, 2007). In Experiment 1, we dissociated the perspectives of the speaker and the addressee, such that whenever the referent was salient for the speaker, it was non-salient for the addressee, and vice versa. Perspective was manipulated by presenting the context sentence directly preceding the speaker's continuation only to the speaker, over headphones, as in Fukumura and Van Gompel (2012). Experiment 2 tested the effect of cognitive load without a perspective difference; that is, all discourse information was shared between speaker and addressee.

5. Experiment 1

5.1. Methods

5.1.1. Participants

Sixty-four students (47 female; mean age, 20.2 years) from Tilburg University participated in the experiment for course credit. Half of them acted as speakers, and the others acted as addressees. All were native speakers of Dutch, the language of the experiment.

5.1.2. Materials

The experimental items consisted of 16 pairs of photographs, taken from Vogels, Krahmer, and Maes (2013), accompanied by two introductory sentences and the onset of a third sentence. The first picture of a pair always showed one male and one female person

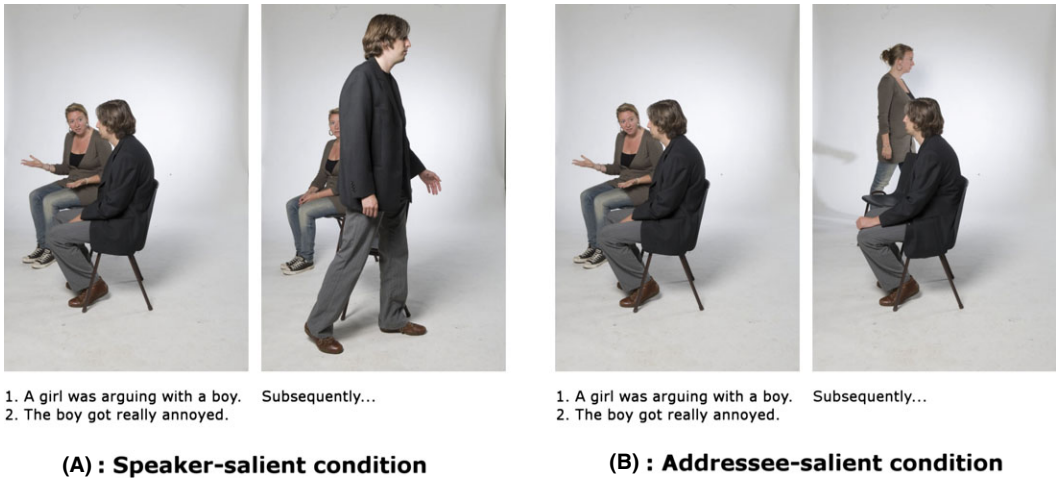

Fig. 1. Example of a stimulus item in two conditions in Experiment 1. Sentence 1 was read aloud by the speaker; sentence 2 was presented only to the speaker, over headphones. Context sentences are translations of the Dutch originals.

sitting next to each other. In the second picture, one of these persons performed an action, such as walking away or getting a glass of water. This person will be referred to as the target character, as participants were expected to refer to this character in their continuations. There were two versions of each picture pair: one in which it was the male person and one in which it was the female person who performed the action. An example of a picture pair is shown in Fig. 1.

The first sentence introduced both characters with indefinite noun phrases, which were either *een meisje* "a girl" and *een jongen* "a boy," or *een vrouw* "a woman" and *een man* "a man." One of these was mentioned as the subject, and the other in a prepositional phrase (e.g., *Een meisje zat te discussiëren met een jongen* "A girl was arguing with a boy"). This sentence was read aloud by the speaker to the addressee. The second sentence described some emotional or physical state of the person mentioned in the prepositional phrase (e.g., *De jongen raakte enorm gepikeerd* "The boy got really annoyed"). Hence, there was always a topic shift between the first and the second context sentence. The second sentence was prerecorded by a female native speaker of Dutch and was only heard by the speaker over headphones. The onset of the third sentence was always *Vervolgens*... "Subsequently...," serving as a cue for the speaker to complete the story based on the second picture. In the speaker-salient condition (condition A in Fig. 1), the target character in the second picture was the subject of the second, privileged sentence, and therefore discourse salient for the speaker but not for the addressee. In the addressee-salient condition (condition B in Fig. 1), the target character was the subject of the first sentence, in which case it was discourse non-salient for the speaker but salient for the addressee, since this sentence was the only context sentence heard by the addressee.

All pictures were counterbalanced for gender of the target character, visual salience,[1] and left-right position of the characters. Twenty additional picture pairs served as fillers.

These differed from the experimental items in that some showed either two male or two female characters (and hence pronouns would be ambiguous) or only one character. In this way, participants were discouraged to use only one type of expression throughout the experiment. In the accompanying sentences, some characters were given labels such as *een verkoopster* "a saleswoman" or *een Duitser* "a German," and sometimes the same character was the subject of both introductory sentences. An additional four items were included as practice items.

5.1.3. Procedure

The experiment took place in a quiet room. Two participants were randomly assigned to the role of speaker and addressee. The participant taking the role of speaker was seated at one end of a table, behind a laptop connected to a PST Serial Response Box. The participant taking the role of addressee was seated at the other end of the table, and was given a booklet containing all different picture pairs and an answer sheet. The experiment was run on the laptop using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA), and was only visible to the speaker. The speaker's task was to complete the stories depicted by the picture pairs in such a way that the addressee could pick out the correct picture pair from the booklet.

Crucially, in one half of the experiment, the speaker received a secondary task (cognitive load condition), while there was no secondary task in the other half (no cognitive load condition). In the no cognitive load condition, each trial started with the item number presented on the screen, accompanied by a 500 ms beep, followed by a fixation cross. Then, the first picture of a pair appeared on the left side of the screen. After 3 s, the first introductory sentence appeared below the picture in a red font. The speaker read this sentence aloud to the addressee. After 5 s, the second sentence was presented to the speaker over the headphones. Next, while the first picture remained visible, the second picture appeared automatically on the right side of the screen, together with the onset of the third sentence, which also appeared below the picture in a red font. At this time, recording started, and the speaker had 6 s to complete the story based on the event shown in the picture, by saying it aloud to the addressee. When this interval had elapsed, recording stopped and the pictures and sentences disappeared. The addressee's task was to select the correct picture pair out of three options from the booklet and mark the correct answer on the sheet. In the experimental items, two of the three options differed only in which character performed the action, making correct reference crucial for the addressee to finish his task successfully. The addressee was instructed to give the speaker a cue when the next trial could be started.

In the cognitive load condition, the appearance of the first picture was preceded by the words BAL or DAL (Dutch for "ball" and "valley," respectively),[2] which was presented in the middle of the screen for 1 s. The same happened at the end of the trial, followed by the question *Was dit woord hetzelfde als het vorige woord? (Ja/Nee)* "Was this word the same as the previous word? (Yes/No)." The speaker then pressed either the green/Yes or the red/No button on the response box. Participants did not receive feedback on their answers.

The participants received instructions both orally and in written form. Speakers were explicitly told that the sentence presented over headphones could not be heard by their

addressee, but that they had to pay attention to it nonetheless, since they would be asked about these sentences after the experiment as an attention check. Indeed, speakers were presented with 10 sentences at the end of the experiment, of which they had to select five that had occurred in the experiment. Speakers were also encouraged to pay attention to the secondary task by way of a prize offer for the participant with the fewest errors. To keep the speaker aware of the addressee's needs, the addressee was allowed to ask the speaker clarification questions if anything remained unclear, but only after the speaker had finished the story.[3] After the experiment, speakers were asked informally about the difficulty of the secondary task.

The experiment was divided into two blocks, of which one contained the secondary task and the other did not, counterbalanced for order. Each block was preceded by two practice items. The experimenter was only present during the instructions and the practice trials. The experiment took about 25 min.

5.1.4. Data coding

We transcribed all speakers' continuations of the third sentence, and we coded all references to the target character (excluding possessives and reflexive pronouns). Since the target referent was referred to only once in the majority of the cases, we only analyzed the first subject reference. We excluded 34 cases in which the first subject did not refer to the target referent. In addition, we excluded seven plural references, three indefinite references, one case in which the sentence presented over the headphones was repeated literally, and one missing case. In addition, there were two cases in which the referring expression was repaired. However, because the repair was of the same type in both cases (e.g., "the man... uh the boy"), we kept these cases. In total, we excluded 46 trials (9.0%). The remaining 466 subject references were coded for the type of referring expression: either full noun phrase or pronoun.

5.1.5. Design and statistical analyses

Crossing the two factors referent salience and cognitive load resulted in a 2 (speaker-salient, addressee-salient) × 2 (cognitive load, no cognitive load) within-participants design. Participants were assigned to one of four lists, each of which contained one version of a given item. The items were presented in a pseudorandom order, with at least one filler item between two consecutive experimental items.

We performed a logit mixed model analysis on the log odds for a pronoun (Jaeger, 2008). Referent salience and cognitive load were included as fixed factors, and participants and items as random factors. The fixed factors were centered to reduce collinearity. We attempted to fit a model with a full random effect structure. In case the model did not converge, we excluded random slopes with the lowest variance (as given by the non-converging model summary; Barr, Levy, Scheepers, & Tily, 2013). From the first converging model, we subsequently excluded random slopes that did not significantly contribute to model fit using log-likelihood ratio tests, with an α-level of .20 (Baayen, Davidson, & Bates, 2008; Barr et al., 2013; Jaeger, 2011). Only the final model will be reported.

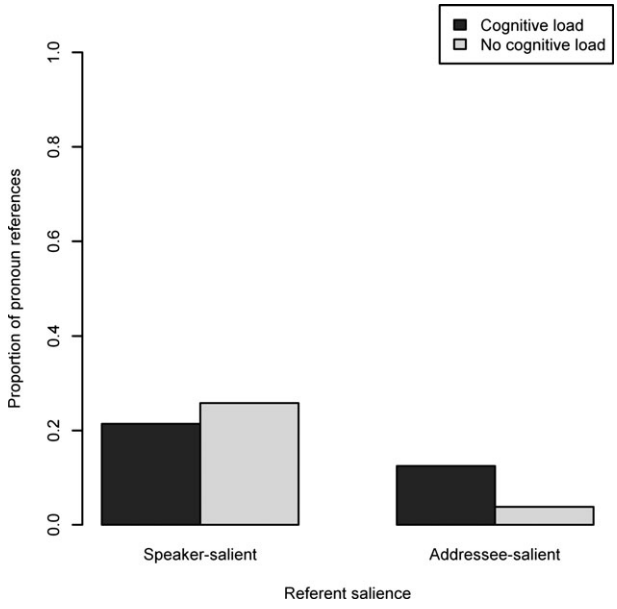

Fig. 2. Proportion of pronoun references to the target character in the four conditions of Experiment 1.

5.2. Results

Fig. 2 shows the proportion of pronoun references to the target character by referent salience and cognitive load condition. The final logit mixed model included random intercepts for participants ($s^2 = 2.87$) and items ($s^2 = 0.15$), as well as by-participant random slopes for referent salience ($s^2 = 12.09$) and cognitive load ($s^2 = 3.04$). We found a main effect of referent salience: Pronouns were more frequent when the referent was discourse salient only for the speaker (23.6%) than when it was discourse salient only for the addressee (8.3%), $\beta = 2.25$; $SE = 0.85$; $p < .01$. There was also a significant main effect of cognitive load, with slightly more pronouns in the cognitive load condition (17.2%) than in the no cognitive load condition (15.8%), $\beta = 1.37$; $SE = 0.56$; $p < .05$. However, these effects were qualified by a significant interaction between referent salience and cognitive load, $\beta = -2.76$; $SE = 0.95$; $p < .01$, suggesting that cognitive load affected pronoun use differently in the two salience conditions.

To arrive at the pairwise comparisons for the interaction effect, we built separate models for the two levels of referent salience using the same procedure as described above. The final model for the speaker-salient condition included a by-participant random slope for cognitive load, while the model for the addressee-salient condition included only by-participant and by-item random intercepts. The effect of cognitive load was not significant in the speaker-salient model, $\beta = 0.77$; $SE = 0.61$; $p = .21$, but marginally so in the addressee-salient model, with pronouns being more frequent in the cognitive load condition (12.5%) than in the no cognitive load condition (3.8%), $\beta = 1.35$; $SE = 0.77$; $p = .08$.

5.3. Discussion

The results of Experiment 1 show that pronouns are more frequent when the referent is discourse salient for the speaker but not for the addressee (i.e., it is the subject of the privileged context sentence) than when the referent is not salient for the speaker but is salient for the addressee. This suggests that even when speakers were not performing a secondary task, pronoun use was dependent more on the referent's accessibility in the speaker's own discourse model than on a calculation of the referent's accessibility in the addressee's discourse model. This is in line with Fukumura and Van Gompel (2012), who found that speakers tend to follow their own discourse model when there is privileged information.

More important, the results suggest that the presence of the secondary task increased the likelihood of pronoun use. Although the effect is small, it is significant, and the interaction between cognitive load and salience suggests that the effect is larger in the addressee-salient than in the speaker-salient condition. This finding is inconsistent with the claim that cognitive load on the part of the speaker decreases the accessibility of the mental representation of the referent in the speaker's discourse model (Arnold & Griffin, 2007). If that were the case, the execution of a dual task should have led to more specific expressions.

Since speakers appeared to be primarily making use of their own discourse model rather than calculating referent accessibility in the addressee's discourse model, the effect of cognitive load seems not to be due to difficulties in perspective taking either. Still, given the relatively low overall proportion of pronouns in Experiment 1, it might be the case that speakers *were* taking into account the addressee's perspective, although not up till the level of calculating the referent's cognitive status for the addressee. This kind of detailed audience design might be cognitively too costly, even without an increased cognitive load (e.g., Bard et al., 2000; Brennan & Hanna, 2009; Horton & Gerrig, 2005). Therefore, speakers may have increased the use of elaborate expressions to be as clear as possible for the addressee, as soon as they were aware of the fact that not all information was shared. This type of audience design might be more difficult under load, which could explain the higher probability of pronoun use in the cognitive load condition.

To determine whether the increase in pronoun use under load in the addressee-salient condition in Experiment 1 is due to the speaker having difficulties in audience design, we conducted a second experiment in which all discourse information was shared between speaker and addressee. If the effect of cognitive load is due to the difference in perspective, changing privileged ground to common ground should cause this effect to diminish or disappear, since there is no need to increase the use of more specific expressions when all information is shared (i.e., speaker's and addressee's discourse models match). If the effect of cognitive load is due to the speaker being more likely to use expressions that are more economical for herself, changing to common ground should not influence this effect, since referent salience remains the same for the speaker.

6. Experiment 2

6.1. Methods

6.1.1. Participants

Sixty-four students (44 female; mean age 22.3 years) from Tilburg University participated in the experiment for course credit. Half of them acted as speakers, and the others acted as addressees. None of them participated in Experiment 1.

6.1.2. Materials

We used the same experimental items as in Experiment 1. The only difference was that the speaker was not wearing headphones and that the second context sentence was presented over the computer speakers. As a result, both speaker and addressee had access to all discourse information.

6.1.3. Procedure

The procedure was identical to that of Experiment 1, except that the speaker was not wearing headphones. As in Experiment 1, the speaker read aloud the first context sentence, which appeared below the first picture. After 5 s the second sentence was presented over the computer speakers. Speakers were told that they had to pay attention to this sentence, since they would be asked about these sentences after the experiment. Next, the speaker completed the third context sentence based on the event shown in the picture. The addressee's task was the same as in Experiment 1, as was the dual task setup in the cognitive load condition. Again, the experiment was divided into two blocks, each preceded by two practice items. After the experiment, speakers were asked to indicate the difficulty of the secondary task on a 7-point Likert scale ("very easy" to "very difficult").

6.1.4. Data coding

The data coding procedure was the same as in Experiment 1. We excluded one case in which the first subject did not refer to the target referent and one plural reference (0.4%).[4] The remaining 510 subject references were coded for the type of referring expression: either full noun phrase or pronoun.

6.1.5. Design and statistical analyses

Crossing the two factors referent salience and cognitive load resulted in a 2 (discourse salient, discourse non-salient) × 2 (cognitive load, no cognitive load) within-participants design. Participants were assigned to one of four lists, each of which contained one version of a given item. The items were presented in a pseudorandom order, with at least one filler item between two consecutive experimental items. Statistical analysis of the data was done in the same way as in Experiment 1.

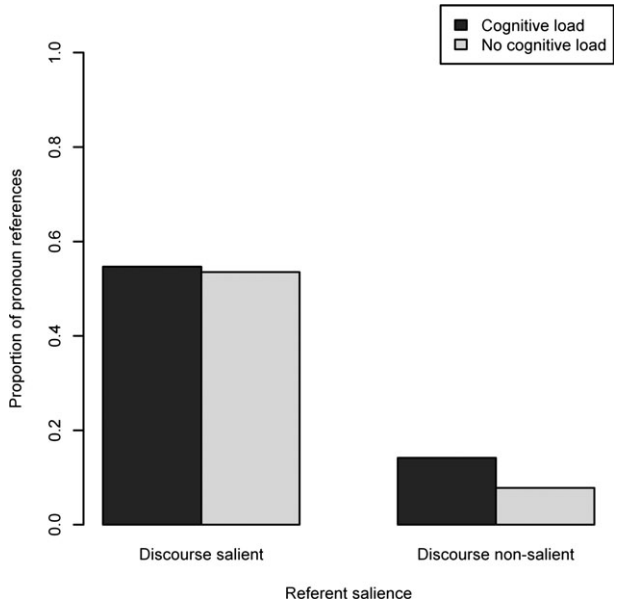

Fig. 3. Proportion of pronoun references to the target character in the four conditions of Experiment 2.

6.2. Results

Fig. 3 shows the proportion of pronoun references to the target character by referent salience and cognitive load condition. The final logit mixed model included random intercepts for participants ($s^2 = 13.51$) and items ($s^2 = 0.26$), and a by-participant random slope for referent salience ($s^2 = 4.52$). We found a main effect of referent salience: Pronouns were more frequent when the referent was discourse salient (54.1%) than when it was not discourse salient (11.0%), $\beta = 6.24$; $SE = 0.88$; $p < .001$. There was also a significant main effect of cognitive load: More pronouns were used when speakers performed a dual task (34.5%) than when they did not (30.6%), $\beta = 0.76$; $SE = 0.39$; $p < .05$. These effects were qualified by a marginally significant interaction, $\beta = -1.34$; $SE = 0.77$; $p = .08$, suggesting that the effect of cognitive load was at least present in the discourse non-salient condition, with pronouns being more frequent in the cognitive load condition (14.2%) than in the no cognitive load condition (7.8%).

6.3. Discussion

Experiment 2 was conducted to investigate whether the effect of cognitive load in Experiment 1 was due to speakers having difficulties in audience design or to speakers producing more economical expressions. The results of Experiment 2 largely correspond to those of Experiment 1, except that speakers seemed to be more likely to produce pronouns, at least in the condition where the referent was salient for the speaker.[5] This suggests that speakers were employing some kind of audience design in the presence of

privileged information, resulting in more specific expressions. The observation that pronouns were more frequent in the cognitive load than in the no cognitive load condition is in line with the finding in Experiment 1 that adding cognitive load increases rather than decreases the probability of pronoun use. Again, this effect seems to be confined to the condition in which the referent is not salient for the speaker. Thus, the results of Experiments 1 and 2 suggest that cognitive load has the same effect when all discourse information is shared as when the salience of the referent differs for the speaker and the addressee.

A tentative comparison of the results of Experiments 1 and 2 indeed suggests that the effect of cognitive load does not differ across the two experiments. Combining the data of both experiments, we built a logit mixed model with referent salience (salient for the speaker, not salient for the speaker), cognitive load (cognitive load, no cognitive load), and experiment (shared context, privileged context) as fixed factors, and participants and items as random factors. Random slopes were only included when they improved model fit (again using log-likelihood ratio tests with an α-level of .20).[6] The final model included random intercepts for participants ($s^2 = 8.73$) and items ($s^2 = 0.30$), as well as by-participant ($s^2 = 18.68$) and by-item ($s^2 = 1.32$) random slopes for referent salience.

There was a significant effect of referent salience, confirming that speakers were more likely to use pronouns when the referent was salient according to their own discourse model (39.0%) than when it was not salient (9.7%), $\beta = 5.53$; $SE = 0.91$; $p < .001$. The effect of cognitive load was also significant, confirming that speakers were more likely to use pronouns when they performed the secondary task (26.2%) than when they did not (23.6%), $\beta = 0.58$; $SE = 0.29$; $p < .05$. There was no main effect of experiment, $\beta = -1.09$; $SE = 0.98$; $p = .27$. However, there was a marginally significant interaction between referent salience and experiment, $\beta = -3.10$; $SE = 1.71$; $p = .07$, suggesting that the effect of salience was smaller in Experiment 1 (privileged context) than in Experiment 2 (shared context). There was also a significant interaction between referent salience and cognitive load, $\beta = -1.51$; $SE = 0.59$; $p < .05$, confirming that the effect of cognitive load differed between the two salience conditions. Most important, the interaction between cognitive load and experiment, as well as the three-way interaction, was not significant, $\beta = -0.24$; $SE = 0.61$; $p = .69$ and $\beta = -0.58$; $SE = 1.25$; $p = .64$, respectively. This indicates that the effect of cognitive load as well as the interaction between cognitive load and referent salience indeed did not differ between the two experiments.

Building separate models for the two levels of referent salience resulted in two models with only by-participant and by-item random intercepts. In the salient for the speaker condition, the effect of experiment was significant: Pronouns were less frequent when there was privileged context (23.6%) than when all discourse context was shared (54.1%), $\beta = -2.62$; $SE = 0.75$; $p < .001$. The effect of cognitive load was not significant, $\beta = -0.15$; $SE = 0.26$; $p = .57$, and neither was the interaction between cognitive load and experiment, $\beta = -0.53$; $SE = 0.53$; $p = .32$. In the not salient for the speaker condition, by contrast, the effect of experiment was not significant, $\beta = 0.41$; $SE = 2.64$; $p = .88$, while the effect of cognitive load was significant, with pronouns being more frequent in the cognitive load condition (13.4%) than in the no cognitive load condition

(6.0%), β = 1.42; *SE* = 0.55; $p < .01$. The interaction between cognitive load and experiment was not significant, β = 0.10; *SE* = 1.21; $p = .94$.

These tentative findings suggest that the presence of privileged information causes the speaker to increase the use of full noun phrases, at least when the referent is salient for the speaker, in which case there is an initial preference for pronouns.[7] The effect of cognitive load seems to be independent of this type of audience design: Under load, speakers prefer to use less costly, more economical referring expressions for non-salient referents, both when there is privileged information and when all information is shared. This effect might therefore be primarily due to the speaker having difficulties in determining that full noun phrases should be used for referents that are not salient in the discourse.

Still, given that the effects of cognitive load were relatively small, and given that perspective (privileged or shared information) was manipulated across different experiments, future studies should investigate the relation between cognitive load and perspective taking in more detail, for example by manipulating both factors in a single, more powerful experiment.

7. General discussion

7.1. Effects of cognitive load

Two experiments investigated the influence of an increased cognitive load for the speaker on the choice of referring expressions. Speakers referred to both salient and non-salient entities (according to their own or their addressee's discourse model), either while performing a secondary task or not. On the basis of the literature, we formulated two alternative hypotheses concerning the impact of cognitive load on referential choice. First, increased cognitive load may result in difficulties for the speaker in taking into account the addressee's perspective. On the one hand, this may increase the speaker's tendency to choose referring expressions based on her own model of the discourse (e.g., Bard & Aylett, 2005; Horton & Keysar, 1996). On the other hand, increased cognitive load may cause speakers to resort to using more economical expressions (i.e., pronouns; Almor et al., 1999; Hendriks et al., 2008, 2014). Second, increased cognitive load may affect the speaker's own discourse model by decreasing the accessibility of referents therein, and hence lead to more elaborate expressions (Arnold & Griffin, 2007; Arnold et al., 2009), irrespective of salience.

The results show that when speakers are under load, they are more likely to produce pronouns in a context for which an accessibility account (e.g., Ariel, 1990) would predict a higher likelihood of full noun phrases. This effect of cognitive load does not seem to be related to the speaker's ability to take the perspective of the addressee, as speakers did not appear to calculate the referent's accessibility for the addressee even when they were not under load: In Experiment 1, as well as in Experiment 2, speakers generally used more pronouns when the referent was salient for them than when it was not salient for them, suggesting that they were basing their choice of referring expressions more on their

own model of the discourse than on assumptions about their addressee's discourse model. Thus, our results do not support the hypothesis that speakers become more likely to fall back on their own discourse model when they experience an increased cognitive load. If that were the case, pronouns in Experiment 1 should have become less frequent under load when the referent was not salient for the speaker but salient for the addressee, and more frequent when the referent was salient for the speaker but not for the addressee.

Although speakers seemed to employ some kind of audience design by increasing the use of full noun phrases when there was privileged information (cf. Fukumura & Van Gompel, 2012), this did not seem to be harmed by the execution of a secondary task: We found a similar effect of cognitive load on the choice of referring expression in Experiment 2, in which speaker and addressee were assumed to have closely aligned discourse models, as in Experiment 1, in which referents that were not salient for the speaker were salient for the addressee.

In addition, our results do not provide support for the hypothesis that an increased cognitive load reduces the accessibility of referents in the speaker's own discourse model, since that would have resulted in an increase of full noun phrases. Hence, the present results suggest that although accessibility may be related to attention, it does not hold generally that less attentive speakers use more elaborate referring expressions.

Our finding that the use of pronouns increases under load is compatible with the hypothesis that cognitive load increases the use of more economical expressions (Hendriks et al., 2014). It is assumed that, in general, speakers prefer economical over elaborate expressions (Almor, 1999; Burzio, 1998; Levinson, 1987), and they may typically only use a full noun phrase when they refer to a character that is not salient in the discourse. Even though speakers do not seem to specifically keep track of the accessibility of the referent for the addressee, their own model of the discourse may serve as a proxy for that of their addressee (e.g., Bard & Aylett, 2005; Dell & Brown, 1991; Pickering & Garrod, 2004). Therefore, using this model to choose referring expressions can still be regarded as a kind of audience design. Thus, assuming that speakers themselves know what they are referring to, the production of full noun phrases may be inherently oriented toward an addressee. In fact, any expression that is more specific, and therefore more costly, than a pronoun could be considered an adaptation to a (hypothetical) addressee (Hendriks et al., 2014). Because this addressee-oriented process may be cognitively effortful, increasing the speaker's cognitive load may reduce this type of audience design. That is, when distracted by a secondary task, speakers may have fewer memory resources available that are needed to infer that a less salient referent should be referred to with a more elaborate expression, such as a full noun phrase. Hence, they are more likely to produce less costly expressions, such as pronouns. For referents that are salient in the discourse, the preference to use a pronoun is already in accordance with the referent's accessibility, which explains why cognitive load does not seem to increase pronoun use for these referents.

Another possible explanation for the increase in pronoun use for non-salient referents is that the speaker has trouble keeping track of the salience of the referent in her own discourse model (as a proxy for her addressee's). In this case, an increased cognitive load may result in less consistent use of referring expressions (Arnold, 2010). That is, pronoun

use may become less tied to the discourse context. For example, speakers under load may use fewer pronouns for referents that are salient in the discourse, and more pronouns for non-salient referents, compared to the normally predicted consistent use of pronouns for salient entities and full noun phrases for non-salient entities. Indeed, Hendriks et al. (2014) found that although elderly adults were capable of taking into account the accessibility of the referent for the addressee, they still used more pronouns for non-salient referents than younger adults. This suggests that they had difficulties keeping track of the salience of the discourse referents. However, such an explanation for our findings does not account for the fact that, in our study, there was no decrease in the use of pronouns for salient referents. After all, if it becomes more difficult to keep track of referent salience, the choice of referring expressions should become less consistent for salient referents as well.

Finally, it could be the case that the effect of cognitive load is at least partly due to the fact that more elaborate expressions are more difficult to produce (i.e., an effect on lexical rather than conceptual representations). For example, with increased memory load, it may be more difficult to retrieve the richer semantic content of full noun phrases from memory, which results in the use of more semantically general expressions such as pronouns (Almor, 1999). This may be less of a problem for salient than for non-salient referents, because the semantic information associated with salient referents might already be more activated, thus accounting for the asymmetry between the salience conditions. To what extent the effect of cognitive load on pronoun use that we found is related to conceptual or lexical representations of referents should be investigated further in future research. See Vogels (2014) for further discussion of how our findings may be incorporated into a broader model of reference production.

7.2. Effects of dissociating the speaker's and addressee's perspectives

As noted above, in addition to the use of full noun phrases for non-salient referents, another type of audience design, emerging from the comparison of the two experiments, seems to be that speakers are more likely to use full noun phrases as soon as it is clear that the preceding discourse is not fully in common ground with their addressee. This might be evidence for a minimal, one-bit model of audience design (Galati & Brennan, 2010; see also Epley, Keysar, Van Boven, & Gilovich, 2004): Speakers use more specific referring expressions as soon as they are aware that not all information is shared, but irrespective of the actual accessibility of the referent for the addressee. Although our findings should be taken with caution, given that perspective was manipulated across different experiments, they are in line with Fukumura and Van Gompel (2012), who found that while speakers were not taking into account their addressee's perspective in choosing referring expressions when the two perspectives were dissociated, they used slightly more pronouns in a condition in which all information was shared (37% vs. 33%), independently of whether the referent was salient or not. This suggests that speakers use more elaborate expressions whenever there is privileged information, even though they might run the risk of being overly specific.

Still, it is striking that the overall proportion of pronouns used in our Experiment 1, which used privileged information, was quite low. In their Experiment 2, for example, Fukumura and Van Gompel (2012) found higher rates of pronoun use in both the privileged, referent-salient condition (48%) and the privileged, referent-non-salient condition (18%) than we did in Experiment 1 (24% and 8%, respectively), while their results for the shared condition (referent-salient: 55%; referent-non-salient: 19%) were similar to our Experiment 2 (54% and 11%, respectively). Part of this difference could be explained by differences in the linguistic materials. For example, while the referent mentioned in the second context sentence was referred to with a pronoun in Fukumura and Van Gompel's experiments, it was referred to with a full NP in our experiments, in accordance with the preferred way of referring to an entity previously mentioned as a direct object in centering theory (e.g., Brennan, 1995). The tendency to pronominalize the entity on a subsequent reference may, however, be stronger when the referent had already been pronominalized. In addition, speakers may have been more likely to reuse the most recent referring expression, which could also have led to more pronouns in Fukumura and Van Gompel's experiments than in ours. However, this difference cannot explain why pronouns were also more frequent when the referent was not salient. Therefore, it seems that speakers in our study were employing minimal audience design more rigorously than in Fukumura and Van Gompel (2012), possibly related to the fact that in their study, the same speakers were presented with both privileged and shared contexts.

One reason why speakers did not seem to make the extra effort to calculate the accessibility of the referent in the addressee's discourse model may be that in the current experiments, as well as in Fukumura and Van Gompel's, references were never ambiguous, since the two characters always had a different gender. Therefore, not taking into account the addressee's perspective would probably not result in interpretation errors. However, when it is clear that not taking into account the addressee's perspective *would* lead to interpretation errors, speakers may be more likely to adapt their choice of referring expressions to the knowledge of their addressee (e.g., Ferreira, Slevc, & Rogers, 2005; Horton & Keysar, 1996). In that case, an increased cognitive load might make this perspective taking more difficult, and cause speakers to fall back on their own perspective.

The filler materials of Experiment 1 contained stories with characters of the same gender, and hence pronouns were ambiguous. Here, we indeed found more pronouns when the referent was salient for the addressee but not for the speaker (17; 33%) than when the referent was salient for the speaker but not for the addressee (7; 13%), suggesting that speakers were taking their addressee's perspective into account. However, cognitive load did not seem to cause speakers to use their own discourse model. Rather, a pattern similar to that in Fig. 2 emerged, with more pronouns under load for referents that were not salient for the speaker. This may be another indication that the effect of cognitive load as manipulated here is independent of perspective taking.

The assumption in the above discussion has been that full noun phrases are more effortful for the speaker than pronouns (Almor, 1999) and are therefore dispreferred unless they would aid the addressee's interpretation. However, it is conceivable, given the frequent overall use of full noun phrases in our experiments, that producing more elaborate referring expressions can also be helpful for the speaker herself. For example,

names for salient entities may be more easily retrieved from the lexicon (Fukumura & Van Gompel, 2012). Alternatively, the use of full noun phrases may be related to the amount of conceptual information that needs to be retrieved to produce a referring expression (Engelhardt et al., 2006; Fukumura & Van Gompel, 2012; Fukumura, Van Gompel, Harley, & Pickering, 2011). Still, it is not clear how such speaker-internal explanations for the use of more elaborate expressions would account for the finding that the use of full noun phrases decreases under load.

Crucially, however, the comparison of the results of our Experiments 1 and 2 tentatively suggest that increased cognitive load does not harm the use of more elaborate expressions as a form of minimal audience design: The effect of cognitive load was the same in the presence of privileged information, in which case speakers often used full noun phrases, as when all information was shared. Hence, using more specific referring expressions when there is no full common ground may be relatively easy. Indeed, it has been argued that this type of audience design is cognitively not very demanding: Information that is available early or takes the form of a binary choice that can be assessed quickly (e.g., *my addressee has heard this or not*) is readily used in choosing referring expressions at little cost (Brennan & Hanna, 2009; Galati & Brennan, 2010). This might also explain the finding by Bard and Aylett (2005) that speakers only seemed to adapt their choice of referring expressions when they switched to a new addressee, since this is information that is easy to take into account (relative to, e.g., building a detailed model of someone else's knowledge). Thus, the decrease in the use of full noun phrases under load in our study does not seem to be related to difficulties in producing elaborate expressions per se, but rather to difficulties in determining *when* a full noun phrase should be used.

7.3. Task-dependencies and individual differences

It is important to note that the effect of cognitive load on referential choices may depend on the specific task and on what part of cognition is actually loaded. For example, in Arnold and Griffin (2007) and Fukumura et al. (2010), the number of referential competitors was varied, which led to divided attention to multiple possible referents. This might be a different kind of cognitive load than divided attention over multiple tasks. Thus, our results do not contest that referents competing for attention affect accessibility and therefore referential choices. However, they are inconsistent with the more general claim that accessibility and hence choice of referring expression is driven by speaker attention (Arnold & Griffin, 2007; Brennan, 1995). Rather, speakers may default to using pronouns when attention is led away from the discourse by an increased memory load.

In a dual task setup, the nature of the secondary task may also make a difference (e.g., whether it is visual or verbal; Baddeley & Hitch, 1974; Kellogg et al., 2007). For example, in the present experiments the use of a verbal secondary task may have especially hindered attention to the linguistic context. It could be the case that other manipulations, such as a visual task (cf. Rosa & Arnold, 2011) or adding time pressure (cf. Horton & Keysar, 1996), interfere more with activating non-linguistic representations or with

perspective taking. In addition, it is conceivable that the artificial nature of our main, referential, task caused some additional load or encouraged the use of task-specific strategies. For example, the modality switches in the context sentences (reading aloud, then listening, then speaking) may have caused an extra increase in cognitive load. The same may hold for the unnatural dissociation of perspective using privileged information presented over headphones. This may therefore have caused a confound between perspective and cognitive load, although that cannot explain the effect of our independent manipulation of cognitive load using a secondary task. These issues need further research.

Finally, our results suggest that there was quite some individual variation as to how speakers' referring expressions were affected by the dual task. Although our cognitive load manipulation had an impact on referential choices, the secondary task appeared to be relatively easy for many participants. Informal inspection of the data suggested that participants who reported to have found the task difficult showed the clearest effects of cognitive load. One cause of individual differences in task difficulty could be the use of strategies for remembering the words BAL and DAL. Over the two experiments, two thirds of all participants reported to have used some kind of mnemonic (e.g., putting up one finger for BAL and two for DAL), although these were not always employed from the beginning. Since it is not yet clear how a heavier load might affect our results, this is a concern that should be taken up by future studies.

8. Conclusion

The study presented in this article has shown that speakers use more pronouns when they experience an increased cognitive load, at least when the referent is not salient for the speaker. We have suggested that this is due to difficulties in determining that a referent that is less salient in the discourse should be referred to with a more specific expression. These difficulties result in the production of more economical forms. We have not found support for the hypothesis that speakers under load are less able to take the addressee's discourse model into account, since speakers only seemed to use a cognitively undemanding form of audience design. Neither do our results support the hypothesis that cognitive load, at least in the form of the dual task used here, decreases the accessibility of referents in the speaker's discourse model. Our results are in line with the view that speakers choose referring expressions based on assumptions about the referent's accessibility for the addressee, but they make these assumptions primarily on the basis of their own model of the discourse. They clearly show that neither purely speaker-oriented nor purely addressee-oriented accounts of reference production can explain speakers' referential choices in discourse.

Acknowledgments

A preliminary analysis of Experiment 1 was presented at the 35th Annual Meeting of the Cognitive Science Society, Berlin 2013, and appeared in the conference proceedings.

Notes

1. These stimuli were originally developed to investigate the influence of visual salience on reference production, hence the visual fore- or backgrounding of one of the characters in Fig. 1. In the present experiment, visual salience was counterbalanced across items; that is, in half of the items the target referent was in the foreground in condition A and in the background in condition B. In the other half it was the other way around. In this way, differences in visual salience should not affect our conditions. In addition, as it turned out, Vogels et al. (2013) did not find an effect of visual salience on the choice of pronouns versus full noun phrases.
2. These stimuli were adopted from Goudbeek and Krahmer (2011), who created a slightly modified version of the stimuli used in Kellogg et al. (2007).
3. Speakers were still able to hear the addressee when wearing headphones.
4. We could speculate about reasons why we had much less missing data than in Experiment 1, but it is clear that the headphone manipulation in that experiment constituted a complicating factor that was not present in Experiment 2. See the discussion in Section 7.
5. Note that "salient for the speaker" also means "not salient for the addressee" in Experiment 1 (privileged context) but "also salient for the addressee" in Experiment 2 (shared context). The reverse holds for "not salient for the speaker."
6. The by-participants random slope for experiment was never included, since this factor was between-participants (i.e., participants in Experiment 1 were different from those in Experiment 2).
7. This is different from Fukumura and Van Gompel (2012), who found no interaction between referent salience and the presence of privileged information.

References

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, *106*(4), 748–765.

Almor, A., Kempler, D., MacDonald, M. C., Andersen, E. S., & Tyler, L. K. (1999). Why do alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's disease. *Brain and Language*, *67*, 202–227.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

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

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187–203.

Arnold, J. E., Bennetto, L., & Diehl, J. J. (2009). Reference production in young speakers with and without autism: Effects of discourse status and processing constraints. *Cognition*, *110*, 131–146.

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.

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*, 390–412.

Baddeley, A. (1986). *Working Memory*. Oxford, England: Clarendon Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*. Vol. *8*. New York: Academic Press.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*, 1–22.

Bard, E. G., & Aylett, M. P. (2005). Referential form, word duration, and modeling the listener in spoken dialogue. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 173–191). Cambridge, MA: MIT Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, *10*(2), 137–167.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*, 274–291.

Burzio, L. (1998). Anaphora and soft constraints. In P. B. D. Fox, P. Hagstrom, M. McGinnis, & D. Pesetsky (Eds.), *Is the best good enough? Optimality and competition in syntax*. Cambridge, MA: MIT.

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener adaptation in language production. In D. Napoli, & J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore festschrift for Lila Gleitman* (pp. 105–129). Hillsdale, NJ: Lawrence Erlbaum Associates.

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*, 554–573.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339.

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, *96*, 263–284.

Fukumura, K., & Van Gompel, R. (2012). Producing pronouns and definite noun phrases: Do speakers use the addressee's discourse model? *Cognitive Science*, *36*(7), 1289–1311.

Fukumura, K., Van Gompel, R. P. G., Harley, T., & Pickering, M. J. (2011). How does similarity-based interference affect the choice of referring expression? *Journal of Memory and Language*, *65*, 331–344.

Fukumura, K., Van Gompel, R., & Pickering, M. J. (2010). The use of visual context during the production of referring expressions. *The Quarterly Journal of Experimental Psychology*, *63*(9), 1700–1715.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*(1), 35–51.

Gann, T. M., & Barr, D. J. (2014). Speaking from experience: Audience design as expert performance. *Language, Cognition and Neuroscience*, *29*(6), 744–760.

Givón, T. (1983). *Topic continuity in discourse*. Amsterdam: John Benjamins.

Goudbeek, M., & Krahmer, E. (2011). Referring under load: Disentangling preference-based and alignment-based content selection processes in referring expression generation. In K. van Deemter, A. Gatt, R. van Gompel, & E. Krahmer (Eds.), *Proceedings of PRE-Cogsci: Bridging the gap between computational, empirical and theoretical approaches to reference*. Boston, 20 July 2011.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics*. Vol. *III* (pp. 41–58), *Speech acts*. New York: Academic Press.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of 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*, 274–307.

Hendriks, P., Englert, C., Wubs, E., & Hoeks, J. (2008). Age differences in adults' use of referring expressions. *Journal of Logic, Language and Information*, *17*, 443–466.

Hendriks, P., Koster, C., & Hoeks, J. C. J. (2014). Referential choice across the lifespan: Why children and elderly adults produce ambiguous pronouns. *Language, Cognition and Neuroscience*, *29*(4), 391–407.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*, 127–142.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91–117.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Jaeger, T. F. (2011). Post to HLP/Jaeger lab blog. Available at http://hlplab.wordpress.com/2011/06/25/more-on-random-slopes. Accessed June 25, 2011.

Kellogg, R. T., Olive, T., & Piolat, A. (2007). Verbal, visual, and spatial working memory in written language production. *Acta Psychologica*, *124*, 382–397.

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.

Levinson, S. C. (1987). Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics*, *23*, 379–434.

Pickering, M., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, *27*, 169–226.

Rosa, E. C., & Arnold, J. E. (2011). The role of attention in choice of referring expression. In K. van Deemter, A. Gatt, R. van Gompel, & E. Krahmer (Eds.), *Proceedings of PRE-Cogsci: Bridging the gap between computational, empirical and theoretical approaches to reference* Boston, 20 July 2011.

Vogels, J. (2014). Referential choices in language production: The role of accessibility. PhD dissertation. The Netherlands: Tilburg University.

Vogels, J., Krahmer, E., & Maes, A. (2013). Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production. *Language and Cognitive Processes*, *28*(9), 1323–1349.

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.

Wubs, E., Hendriks, P., Hoeks, J., & Koster, C. (2009). Tell me a story! Children's capacity for topic shift. In J. Crawford, K. Otaki, and M. Takahashi (Eds.), *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 3)* (pp. 313–324). Somerville, MA: Cascadilla Press.