Producing Pronouns and Definite Noun Phrases: Do Speakers Use the Addressee’s Discourse Model?

Kumiko Fukumura,a Roger P.G. van Gompelb

aSchool of Psychological Sciences and Health, University of Strathclyde
bSchool of Psychology, University of Dundee

Received 25 December 2010; received in revised form 7 December 2011; accepted 10 January 2012

Abstract

We report two experiments that investigated the widely held assumption that speakers use the addressee’s discourse model when choosing referring expressions (e.g., Ariel, 1990; Chafe, 1994; Givón, 1983; Prince, 1985), by manipulating whether the addressee could hear the immediately preceding linguistic context. Experiment 1 showed that speakers increased pronoun use (and decreased noun phrase use) when the referent was mentioned in the immediately preceding sentence compared to when it was not, even though the addressee did not hear the preceding sentence, indicating that speakers used their own, privileged discourse model when choosing referring expressions. The same pattern of results was found in Experiment 2. Speakers produced more pronouns when the immediately preceding sentence mentioned the referent than when it mentioned a referential competitor, regardless of whether the sentence was shared with their addressee. Thus, we conclude that choice of referring expression is determined by the referent’s accessibility in the speaker’s own discourse model rather than the addressee’s.

Keywords: Reference; Language production; Pronoun; Definite noun phrase; Accessibility; Discourse; Common ground; Audience design

1. Introduction

One of the most controversial issues in recent psycholinguistic research on language production is to what extent speakers model the needs of their addressee when choosing particular linguistic forms (e.g., Allbritton, McKoon, & Ratcliff, 1996; Bard & Aylett, 2005; Ferreira & Dell, 2000; Horton & Keysar, 1996; Kraljic & Brennan, 2005; Schafer,
Speer, Warren, & White, 2000; Snedeker & Trueswell, 2003). Clearly, for communication to succeed, speakers must produce utterances that are comprehensible for their addressee. In his seminal work on conversational maxims, Grice (1975) argued that speakers should tailor their utterances so that they are maximally helpful by providing as much information as necessary to avoid ambiguity for their addressee while at the same time not being overly informative. That is, speakers must be cooperative, not egocentric, by designing utterances that are easy for their addressee. The view that language production is a cooperative process, which is often referred to as the audience design hypothesis, was further elaborated by Clark and his colleagues (e.g., Clark, 1996; Clark & Carlson, 1982; Clark & Marshall, 1981; Clark & Murphy, 1982), who argued that speakers formulate utterances by consulting information that is mutually shared with their partner, or common ground. According to Clark and Marshall (1981), speakers keep track of what information is in the common ground with a particular addressee by constructing a detailed model of the addressee’s knowledge and beliefs and constantly updating it. Furthermore, the audience design hypothesis has also strongly influenced the way computational linguists have modeled the generation of referring expressions, starting with early work by Appelt (1985a,b; Appelt & Kronfeld, 1987), who assumed that speakers choose referring expressions on the basis of mutual knowledge with the addressee. Gricean maxims also play a crucial role in Dale and Reiter’s (1995) influential Incremental Algorithm, which is based on the idea that referring expressions should be sensitive to the needs and abilities of the addressee (see also Dale, 1992). Following these earlier models, many later computational algorithms also aim to produce expressions based on the addressee’s perspective (e.g., Heeman & Hirst, 1995; Koller & Stone, 2007; O’Donnell, Cheng, & Hitzeman, 1998). However, the extent to which the model of the addressee influences speakers’ linguistic performance has been under much debate in psycholinguistics (e.g., Dell & Brown, 1991; Horton & Gerrig, 2005a, b; Horton & Keysar, 1996; Pickering & Garrod, 2004). Some of the linguistic phenomena that have been ascribed to audience design have been explained in terms of production-internal constraints that favor speech that is easy to produce (e.g., Brown & Dell, 1987; Pickering & Garrod, 2004).

Much evidence suggests that speakers take into account the addressee’s needs when producing referring expressions. For instance, studies have found that speakers are more likely to avoid ambiguous bare noun phrases (e.g., the circle) when a referential competitor of the same semantic category (another circle) is visible to their addressee than when it is not (e.g., Horton & Keysar, 1996; Nadig & Sedivy, 2002), indicating that speakers can use the addressee’s visual perspective when avoiding referential ambiguity. Similarly, Matthews, Lieven, Theakston, and Tomasello (2006) found that even children as young as 3 and 4 years old take into account whether the addressee can see the referent when using pronouns that have no linguistic antecedent in the immediate context (i.e., unheralded pronouns, Gerrig, 1986). Other studies have found that speakers tend to avoid expressions that require expert knowledge that is not shared with the addressee (Isaacs & Clark, 1987) and expressions that are based on conceptualizations formed with a different addressee and that may therefore be unfamiliar to the current addressee (Brennan & Clark, 1996; Horton & Gerrig, 2002, 2005b; Wilkes-Gibbs & Clark, 1992).
All these studies focused on cases where taking the perspective of the addressee was essential for avoiding ambiguity or potential referential failure. In Horton and Keysar (1996) and Nadig and Sedivy (2002), using a bare noun phrase was ambiguous when the addressee could see a referential competitor of the same category, but it was unambiguous when the addressee could not see the competitor. Similarly, in Matthews et al. (2006), unheralded pronouns did not identify who the referent was when the addressee could not see the referent. In other studies, reference would have been ambiguous to the addressee if she or he did not share the knowledge presupposed by the referring expression (e.g., Horton & Gerrig, 2002, 2005b; Isaacs & Clark, 1987; Wilkes-Gibbs & Clark, 1992). This raises the question of whether audience design is largely driven by the speaker’s communicative effort to avoid otherwise uncommunicative referring expressions.

Indeed, it has been argued that audience design primarily occurs in utterance repairs (Barr & Keysar, 2006; Brown & Dell, 1987; Dell & Brown, 1991; Horton & Keysar, 1996): Routine language production processes proceed egocentrically, but during later production processes, speakers monitor and adjust the infelicity of their utterance by accommodating the addressee’s perspective (the monitoring and adjustment hypothesis). According to this idea, the speaker’s knowledge of what the addressee knows or does not know only optionally influences language production processes (see Barr & Keysar, 2006 for a review). In keeping with this, the results of Horton and Keysar (1996) suggest that visual perspective-taking only occurs when speakers are not under time pressure: When they are under time pressure, they tend to produce as many modifiers when the addressee can see the competitor as when s/he cannot. Similarly, Rossnagel (2000, 2004) found that if speakers are under high cognitive load when instructing the assembly of a machine, they often fail to take into account whether they are speaking to a student or a 7-year-old boy and fail to respond to the addressee’s feedback as effectively as when they are under low load. This also suggests that perspective adjustment is subject to other cognitive constraints. Furthermore, Jucks, Becker, and Bromme (2008) showed that when no feedback from the addressee is available (e.g., responding to email enquiries), experts fail to take into account the addressee’s expertise.

According to many theories of reference (Ariel, 1990; Chafe, 1976, 1994; Clark & Marshall, 1981; Givón, 1983, 1992; Grosz, Joshi, & Weinstein, 1995; Gundel, Hedberg, & Zacharski, 1993; Marslen-Wilson, Levy, & Tyler, 1982; Prince, 1985), however, audience design goes beyond avoiding ambiguous or outright inadequate reference but affects the choice of referring expressions more generally. The key assumption of these theories is that speakers choose particular forms of reference depending on how accessible or salient the referent is for the addressee. That is, speakers choose less explicit referring expressions such as pronouns when they assume the referent is highly salient in the addressee’s discourse model, whereas more explicit referring expressions such as names or definite noun phrases are favored when they assume that the referent is less salient in the addressee’s model. For instance, Chafe (1976) argued that ‘‘Given (or old) information is that knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance. So-called new information is what the speaker assumes he is introducing into the addressee’s consciousness by what he says’’ (p. 30) and suggested that a pronoun is used when speakers assume that the referent is ‘‘in the addressee’s consciousness’’ (p. 31).
These theories commonly assume that speakers use the form of reference as a processing signal to their addressee. By using less explicit referring expressions, speakers signal that the referent is highly accessible in the addressee’s model, whereas by using more explicit referring expressions they signal that addressees should retrieve an entity that is less accessible in the addressee’s memory. For instance, Givón (1992) argued that speakers use referring expressions “to accommodate the hearer’s perspective” and “to ground the information into the hearer’s perspective” (p. 8). Similarly, Ariel (2001) claimed that “referring expressions instruct the addressee to retrieve a certain piece of given information from his memory by indicating to him how accessible this piece of information is to him at the current stage of discourse” (p. 29).

Perhaps surprisingly, however, there has been very little evidence that speakers choose referring expressions based on the accessibility or salience of the referent in the addressee’s discourse model (rather than their own). Previous studies have shown that reduced referring expressions such as pronouns (relative to more explicit referring expressions such as names or definite noun phrases) are more frequent when the referent is more frequently mentioned in the prior discourse (Ariel, 1990; Givón, 1983) or mentioned in the more recent linguistic context (Ariel, 1990; Givón, 1983). Other studies have found that the use of pronouns is affected by the antecedent’s structural position in the preceding sentence (e.g., Arnold, 2001; Brennan, 1995; Fletcher, 1984; Fukumura & Van Gompel, 2010; Stevenson, Crawley, & Kleinman, 1994), the animacy of the referent (Fukumura & Van Gompel, 2011), the presence of a referential competitor in the preceding linguistic context (Arnold & Griffin, 2007) and in the visual context (Fukumura, Van Gompel, & Pickering, 2010), and the degree of similarity between discourse entities (Fukumura, Van Gompel, Harley, & Pickering, 2011). However, these findings may be due to the referent’s accessibility in the language user’s own model rather than in the addressee’s model; that is, speakers may be more likely to use pronouns rather than names or definite noun phrases when the referent is highly accessible in their own discourse model rather than in their addressee’s model.

To determine whether the choice between pronouns and definite noun phrases is affected by the addressee’s discourse model, the current study systematically manipulated the referent’s accessibility to the speaker and to the addressee. During conversations, interlocutors do not always share the same discourse representation. For example, conversations sometimes get interrupted by a sudden telephone call or a text message, or by a new interlocutor, who is unaware of the preceding conversation. An important question is whether speakers keep track of whether the discourse representation is shared with their addressee and whether they take this information into account when they choose referring expressions. One study that addressed this question is Bard and Aylett (2005). In their study, speakers had to describe a map twice to two different addressees. Speakers reduced the articulation of references to landmarks when describing the map to the second addressee compared to the first addressee, suggesting that a reduction in articulation occurred regardless of whether the addressee heard prior reference to it (see also Bard, Lowe, & Altmann, 1989). In contrast, speakers did not significantly attenuate their form of reference when describing the map to the second addressee. This led Bard and Aylett to propose the dual-process hypothesis: Articulatory planning, which is driven by fast, automatic processes, is not affected by
the addressee’s knowledge, whereas the form of the expression, which involves slower, more controlled production processes, is. More recently, Galati and Brennan (2010) also suggested that the specificity of referring expressions is affected by the addressee’s knowledge. In their study, speakers produced less detailed, and less intelligible descriptions when they retold a story to the same addressee compared to when they retold a story to a different addressee, whereas the durations of the descriptions were not affected by who speakers spoke to.

However, it remains unclear whether speakers choose referential forms by taking into account the accessibility of a specific referent for their addressee as the discourse progresses. Whether one has already described the same story to a particular addressee is clear from the addressee’s identity, and the accommodation of such information may represent a very coarse form of addressee adaptation, as argued by Galati and Brennan (2010): ‘‘All it takes … would be a single either/or cue that could make it extremely easy for speakers to track an audience’s needs … a one-bit, most minimal partner model’’ (p. 47). Indeed, Van der Wege (2009) reported that the specificity of the speaker’s descriptions varies depending on who speakers think they are speaking to. In her study, speakers produced more detailed descriptions when they were asked to describe pictures for an imagined addressee than when they had a real addressee, and speakers were least specific when they were given no explicit addressee. What is unclear, though, is whether speakers choose referential forms based on the referent’s accessibility in the addressee’s discourse model. Keeping track of the addressee’s discourse model on a moment-by-moment basis when choosing referring expressions may be a fairly complex process. It is therefore possible that, contrary to the predictions of many discourse theories, speakers do not engage in such activity during conversations.

Thus, we conducted two referential communication experiments to examine whether the addressee’s discourse model affects the speaker’s choice of anaphoric expressions. Specifically, we investigated if speakers choose pronouns or repeated noun phrases depending on whether the addressee also heard an immediately preceding sentence that affected the referent’s accessibility to the speaker. If speakers take the addressee’s discourse model into account, we expected that speakers should produce more pronouns when the addressee heard the immediately preceding sentence that made the target referent more accessible than when the addressee did not hear this sentence. In contrast, if speakers primarily choose referring expressions on the basis of their own discourse model, they should produce more pronouns when the immediately preceding sentence makes the target referent more accessible, even when their addressee did not hear the sentence. For both experiments, we tested pairs of participants in a referential communication task, in which one participant took part as a speaker and the other as an addressee, and examined the speaker’s choice of referring expressions while s/he described pictures to direct the addressee, who had to recreate the scene viewed by the speaker. As explained below, the utterances that preceded the target descriptions were tightly controlled, which enabled us to systematically vary the referent’s accessibility to the speaker and to the addressee. To ensure that the speaker considered the addressee a true conversational partner (Lockridge & Brennan, 2002), both participants were naïve to the experimental manipulations.
2. Experiment 1

We investigated whether speakers take into account the referent’s accessibility in their addressee’s discourse model when choosing referring expressions. We did this by examining whether speakers’ tendency to use pronouns rather than definite noun phrases is affected by whether the addressee heard the immediately preceding sentence or not. If the choice of referring expression is primarily affected by the addressee’s discourse model, then whether the preceding utterance increased the referent’s accessibility should have an effect only if the addressee heard the sentence. In contrast, if speakers primarily choose referring expressions based on their own model, then they should choose referring expressions irrespective of whether the addressee heard the preceding sentence or not. It is also possible that speakers are affected by both: Speakers may take into account the addressee’s discourse model as well as their own, privileged discourse model.

In each trial, both the speaker and the addressee saw a picture of toy characters such as the top panel of Fig. 1 on their own computer screen. Then the addressee laid out the toys on the table and the speaker read aloud the first context sentence (1), which appeared below the picture on the speaker’s computer screen. The sentence was read aloud rather than presented auditorily so that speakers themselves had to introduce the relevant characters (the mermaid and the admiral) linguistically to their partner. Next, speakers heard a pre-recorded second context sentence (2a-b). Sentence (2a) contained a pronoun that referred to the last-mentioned character in the first sentence (1) (the admiral), whereas sentence (2b) contained a pronoun referring to the first-mentioned character in (1) (the mermaid). As described in detail below, we manipulated the presentation of this second context sentence such that it was either presented via headphones to the speaker so that only the speaker could hear the second context sentence (privileged context) or presented via loudspeakers so that both the speaker and the addressee could hear it (shared context). Following the auditorily presented sentence, speakers saw a second picture (the bottom panel of Fig. 1), which showed an action carried out by the last-mentioned character in sentence (1) (the admiral, hereafter the target). The speaker had to describe the action and the addressee, who could not see the picture, acted out the speaker’s description using the toys. We examined the proportion of pronouns relative to repeated noun phrases in productions of the target utterance (e.g., saying He/The admiral stands up). The first-mentioned character in sentence (1) (the mermaid, hereafter the competitor) always had a different gender than the target, so reference was unambiguous regardless of whether speakers produced a definite noun phrase or pronoun.

1. The mermaid is waiting for a taxi with the admiral.
2. a. He is sitting in a wheelchair.
   b. She is sitting on a bench.

We tested three conditions in Experiment 1. The first condition was the target mentioned-shared condition. In this condition, the second context sentence referred to the target character (2a), thus making the target more salient, and the sentence was presented via loudspeakers,
which established a shared context because both the speaker and the addressee heard the sentence. In contrast, in the second condition, the \textit{target mentioned-privileged condition}, the second context sentence (2a) also referred to the target, but it established a privileged context, because it was presented via headphones that the speaker was wearing, so the addressee could not hear the sentence. Finally, in the third condition, the \textit{competitor mentioned-privileged condition}, the second context sentence (2b) was also presented via headphones to the speaker, but it referred to the competitor (the mermaid), making the competitor more salient. In order to avoid possible confusion for the speaker as to whether the addressee heard the critical second context sentence, each condition was tested in a separate block (counterbalanced for order across the experiment), but the same participant played either the speaker or addressee role in all conditions. We also made sure that the speaker (and the addressee in the shared condition) listened to the second sentence for comprehension by having them verify that the sentence corresponded to the first picture.
The referent should be more accessible to the addressee when s/he had just heard it referred to (target mentioned-shared) than when s/he had not (target mentioned-privileged). Thus, if speakers adjust to the referent’s accessibility in the addressee’s model, they should produce more pronouns (therefore, fewer repeated noun phrases) in the target mentioned-shared than the target mentioned-privileged condition. In contrast, if speakers use their own discourse model, it should not matter whether the addressee heard the reference to the target in sentence (2a) or not, so the target mentioned-privileged and target mentioned-shared conditions should not differ. Furthermore, both conditions should result in more pronouns compared to the competitor mentioned-privileged condition, where the immediately preceding sentence (2b) refers to the competitor, which should make the referent less accessible to the speaker.

2.1. Method

2.1.1. Participants
Twenty-four pairs of participants from the University of Dundee who were native speakers of British English (aged 17–30) took part in return for payment or course credit. None of them reported to be dyslexic (which was important because they had to read the first context sentence).

2.1.2. Materials
We constructed 24 experimental item sets. Each item set consisted of two photographs of miniature toy characters (such as a king, a queen, a pirate, or a mermaid), a written sentence and an auditory sentence. Fig. 1 presents an example photograph panel. The top half of each panel introduced two toy characters of different gender (the target and the competitor), and the bottom half depicted a simple action carried out by the target (e.g., standing up from a wheelchair). The positions of the target and competitor characters were counterbalanced between items.

Both the target and the competitor were linguistically introduced in a written sentence, as in (1), where the target (e.g., the admiral) was introduced as the prepositional object in a with-phrase and the competitor (e.g., the mermaid) as the subject. For each item, we created two auditory sentences, where either the target (the admiral) or the competitor (the mermaid) was mentioned, thereby making that character prominent. In the target-mentioned condition (2a), the sentence began with a pronoun referring to the target (the admiral), and in the competitor-mentioned condition (2b), it began with a pronoun referring to the competitor character. We used a pronoun rather than a repeated noun phrase in both cases, because although pronouns tend to be more frequently produced when the antecedent is a syntactic subject than a nonsubject (e.g., Arnold, 2001; Fukumura & Van Gompel, 2010; Stevenson et al., 1994), past research has shown that pronouns are common even for nonsubject antecedents. When pronouns were unambiguous (as in the current study), participants produced pronouns to refer to the object in the preceding sentence 83% of the time in Fukumura and Van Gompel (2010), and 77% of the time in Fukumura and Van Gompel (2011), and participants in Fukumura et al.’s (2010) study produced pronouns 50% of the time when referring to a possessive modifier of the subject. Furthermore, Gundel, Grosz, and Gilliom (1993)
found that reading times for sentences with pronouns referring to a nonsubject were not significantly slower than reading times for pronouns referring to a subject.

In half of the items, the sentence correctly described the picture, whereas in the other half, it did not. The sentence was recorded at normal speaking rate by a female native speaker of British English, sampled at 22 kHz. The mean durations for the referent-mentioned condition (1.62 sec) and the competitor-mentioned condition (1.67 sec) did not differ significantly, $t(23) = 1.30$, $p = .208$. Pronouns were pronounced equally clearly in the target-mentioned and competitor-mentioned conditions. In addition, 12 practice and 36 filler items were constructed.

2.1.3. Procedure and design

Before the experiment, the pair of participants was told that the experiment was about how people communicate verbally when they cannot see each other. They drew lots to determine who was the speaker and who was the addressee. The speaker and the addressee were then seated side by side at a table, facing a computer screen, and a board between them prevented them from seeing each other. The experimenter then explained the tasks orally by presenting the stimuli from the first practice trial. The visual stimuli (the photographs and a context sentence) were presented using DMDX software (Forster & Forster, 2003) and the auditory stimuli were presented either over loudspeakers to both participants (shared condition) or over headphones to the speaker alone (privileged condition). The speaker’s speech was recorded on a MiniDisk, which was later used for coding.

At the beginning of each trial, both the speaker and the addressee saw a photograph of miniature toy characters on their screen. The addressee received the toys from the experimenter and recreated the scene depicted in the photograph on the table, so that the speaker sitting on the other side of the board could also see the toys. Once the objects were laid out, the speaker pressed a key, which triggered the presentation of a written sentence below the first photograph on the speaker’s computer screen (the addressee did not see this sentence or the following photograph). The speaker read aloud the context sentence to introduce the characters to the addressee, and s/he then pressed a key, which prompted the presentation of a pre-recorded auditory sentence (2a & 2b). In the target mentioned-shared condition, sentence (2a) was presented via the loudspeakers on the table that participants sat at, and both the speaker and the addressee judged whether the sentence was consistent with the photographs, by pressing a yes or no button. This was done to ensure that the speaker would pay attention to the auditorily presented context sentence. In contrast, in the privileged condition, speakers heard sentence (2a) (target-mentioned-privileged condition) or (2b) (competitor-mentioned-privileged) over headphones and only the speaker judged whether the sentence matched the picture. Next, a second photograph appeared below the first picture on the speaker’s screen, replacing the context sentence. The picture depicted an action carried out by the target character, which speakers had to describe to the addressee. The addressee could not see the picture and had to act out the description using the toys. The speaker then indicated whether the addressee’s action corresponded to the one in the photograph by pressing the yes or no button.

Thus, there were three conditions in total: target mentioned-shared, target mentioned-privileged, and competitor mentioned-privileged conditions. The conditions were presented in three
separate blocks, and the order of blocks was rotated in six permutations, which comprised six lists, each of which contained 24 experimental items and 36 filler items. Each list had eight experimental items from each condition, with one version of each item occurring in each list, presented in a fixed quasi-random order, subject to the constraint that the same character did not occur consecutively. Four pairs of participants were randomly assigned to each list. To ensure that speakers were fully aware when their addressee could hear the preceding sentence or not, participants were asked to swap tasks during the practice sessions of each block (i.e., speakers had to play the addressee’s role) and the experimenter reiterated that when the sentence was presented via the headphones, the addressee could not hear the sentence. There were four practice trials before the start of each block. The experiment took around 45 min.

2.1.4. Scoring
We scored whether participants produced a pronoun or a repeated noun phrase in cases where they referred to the referent character as the subject in the first sentence they produced. We excluded trials where participants did not refer to the referent character (6 trials); they used a different noun phrase instead of a repeated noun phrase (such as the boy rather than the prince) (18 trials) or dropped the subject (1 trial). In total, 25 trials (4.3% of responses) were excluded.

2.2. Results
Fig. 2 presents the mean percentages of pronouns out of all pronoun and repeated noun phrase responses by condition. We conducted two ANOVAs on arcsine-transformed proportions of pronouns (Winer, 1971), one on the participant means with participants as the random variable (F1) and one on the item means with items as the random variable (F2). Condition was treated as a within-participants and within-items variable and we also included participant list (I-VI) as a between-participants variable in the participant analysis and item list (I-VI) as a between-items variable in the item analysis in order to eliminate variance caused by random differences between groups (Pollatsek & Well, 1995). The anal-

Fig. 2. Mean percentages of pronouns out of all pronouns and repeated noun phrases in Experiment 1. Bars represent standard errors.
yses revealed a main effect of condition, $F(1, 36) = 12.64, p < .001, \eta^2_p = .412$; $F(2, 42) = 25.34, p < .001, \eta^2_p = .547$. Planned comparisons showed that pronouns were less frequent in the competitor mentioned-privileged (16%) than in the target mentioned-privileged condition (37%), $F(1, 18) = 18.22, p < .001, \eta^2_p = .503$; $F(2, 21) = 21.90, p < .001, \eta^2_p = .511$. Although pronouns were slightly more frequent in the target mentioned-shared (44%) than the target mentioned-privileged condition (37%), this difference did not reach significance, $F(1, 18) = 3.22, p = .087, \eta^2_p = .133$.

2.3. Discussion

Speakers produced more pronouns in the target mentioned-privileged than the competitor mentioned-privileged condition even though their addressee did not hear the second sentence in either condition. This suggests that speakers based their choice of referring expression on the referent’s accessibility in their own discourse model rather than on the addressee’s model. Furthermore, the difference between the target mentioned-shared and target mentioned-privileged conditions was not significant, consistent with the idea that speakers relied on their own, privileged context.

3. Experiment 2

Although the target mentioned-shared and the target mentioned-privileged conditions did not significantly differ in Experiment 1, the direction of the means might suggest that speakers took into account the fact that the target character was more accessible to the addressee when the addressee could hear reference to the target in sentence (2a) compared to when s/he could not. However, the nonsignificant difference may have occurred not because speakers took into account the accessibility of the target for their addressee, but because regardless of whether the second sentence referred to the target character (and therefore made it more accessible) or not, they were somewhat more explicit when the addressee did not share the same context.

We therefore conducted Experiment 2 by adding a new condition, where the immediately preceding sentence mentioning the competitor was presented by loudspeakers (competitor mentioned-shared condition), so both the speaker and the addressee heard reference to the competitor in the second context sentence. That is, we orthogonally manipulated (a) whether the referent or competitor was mentioned in the immediately preceding sentence (2a vs. 2b) and (b) whether this sentence was shared with the addressee or not (experimental block with loudspeakers vs. block with headphones). The context sentence (1) and the photographs (Fig. 1) were the same as in Experiment 1. If speakers generally use fewer pronouns when their addressee did not hear the immediately preceding sentence compared to when s/he did (i.e., they are generally more explicit when their addressee did not hear the sentence), we expect that pronouns should be more frequent not only when their addressee heard the reference to the target but also when their addressee heard the reference to the competitor.
compared to when s/he did not hear the sentence. If so, there should be a main effect of sharedness. In contrast, if speakers do take into account the referent’s accessibility in the addressee’s model, we expect more pronouns when the addressee heard reference to the target than when only the speaker heard reference to the target, but pronouns should be less frequent when the addressee heard reference to the competitor than when only the speaker heard this. This should result in an interaction between reference in the second context sentence (target mentioned, 2a vs. competitor mentioned, 2b) and sharedness (shared vs. privileged).

3.1. Method

3.1.1. Participants
   Thirty-two pairs of participants from the same population as in Experiment 1 took part. None of them had participated in the previous experiment.

3.1.2. Materials
   We used the same 24 experimental items as in Experiment 1.

3.1.3. Procedure and design
   These were the same as in Experiment 1, except for the following amendments. We created an additional condition in which the sentence mentioning the competitor (2b) was presented via loudspeakers. This resulted in a $2 \times 2$ repeated measures design: second context sentence (referent mentioned vs. competitor mentioned) $\times$ sharedness (shared vs. privileged). Together with the 36 filler items, the 24 items were distributed across four lists, each containing six items from each condition, and one version of each item. Sharedness was manipulated in blocks, and we counterbalanced the order of the blocks as a between-participants and within-items variable. Thirty-two pairs of participants were randomly assigned to four lists, each containing 12 practice trials.

3.1.4. Scoring
   Scoring was done in the same way as in Experiment 1. We excluded one trial due to a technical error and two trials in which addressees inadvertently manipulated the objects in response to the first sentence before the speaker produced the target description. We also excluded one trial in which a participant referred to both characters as they and 15 trials in which participants used a different noun phrase instead of a repeated noun phrase. In total, 19 trials (2.5% of all responses) were excluded.

3.2. Results

Fig. 3 presents the means. We conducted ANOVAs on the arcsine-transformed proportions of pronoun responses with reference in second context sentence and sharedness as
within-participants and within-items variables and participant/item list (I–IV) as a between-participants and between-items variable.

The analyses revealed a main effect of second sentence reference, $F_1(1, 28) = 46.28, p < .001, \eta^2_p = .623$, $F_2(1, 20) = 148.51, p < .001, \eta^2_p = .881$, indicating that participants used more pronouns when the preceding sentence mentioned the referent (52%) than when it mentioned the competitor (19%). The main effect of sharedness was not significant by participants, $F_1(1, 28) = 1.98, p = .171, \eta^2_p = .066$, but very close to significance by items, $F_2(1, 20) = 4.19, p = .054, \eta^2_p = 173$, which indicated a tendency for fewer pronouns in the privileged (33%) than in the shared (37%) condition. Importantly, there was no significant interaction between reference in the second sentence and sharedness, $Fs < 1$.

Planned comparisons revealed that the effect of context sentence was significant in both the shared conditions, $F_1(1, 28) = 42.94, p < .001, \eta^2_p = .605; F_2(1, 20) = 70.14, p < .001, \eta^2_p = .778$, with fewer pronouns in the competitor-mentioned (19%) than target-mentioned condition (55%), and privileged conditions, $F_1(1, 28) = 31.30, p < .001, \eta^2_p = .528; F_2(1, 20) = 99.93, p < .001, \eta^2_p = .833$, again with fewer pronouns in the competitor-mentioned (18%) than target-mentioned (48%) condition.

3.3. Discussion

Pronouns were more frequent when the second sentence mentioned the referent (2a) than when it mentioned the competitor (2b), indicating that the referent was more accessible when it was mentioned in the immediately preceding sentence than when the competitor was mentioned. Crucially, this effect was not modulated by whether the sentence was shared or privileged, and speakers significantly increased pronoun use when the immediately preceding sentence mentioned the referent rather than the competitor even if that sentence was not shared with their addressee. In addition, the by-item analyses showed a tendency for fewer pronouns when the addressee did not listen to the second sentence than when s/he
did, regardless of whether the sentence made the referent or the competitor more accessible. This may represent what Galati and Brennan (2010) called “a one-bit, most minimal audience design,” because all speakers need to know is whether the addressee heard the preceding sentence. Importantly, such audience adaptation does not require speakers to model how accessible the referent is in the addressee’s discourse model: to compute how accessible the referent is to the addressee, they need to know not only whether the addressee heard the preceding sentence but also what entity was mentioned in that sentence. That is, the referent is more accessible to the addressee when s/he heard the sentence that mentioned the referent than when s/he did not, but the referent is less accessible when the addressee heard the sentence that mentioned the competitor than when s/he did not. Given that speakers clearly knew which character was mentioned in the preceding sentence (fewer pronouns in the competitor-mentioned than the referent-mentioned condition), the lack of addressee accessibility effect was most likely due to speakers failing to combine multiple sources of information.

As discussed earlier, past research has shown that speakers frequently produce pronouns both when the referent of the pronoun is the syntactic subject in the preceding utterance (as in the competitor-mentioned second context sentence, 2b) and when it is not (target-mentioned second sentence, 2a). However, the use of a pronoun (rather than a definite noun phrase) in the second context sentence may have been less preferred in the target-mentioned than competitor-mentioned conditions. One might therefore wonder whether this somehow affected speakers’ audience design. First we should note that our main interest was not in the difference between the target-mentioned and competitor-mentioned conditions, but in the interaction with sharedness: If the addressee’s discourse model affects the speaker’s referential choice, the difference between the target-mentioned and competitor-mentioned conditions should be smaller when the addressee did not hear the second sentence than when s/he did. This prediction holds regardless of how preferred the pronoun is in the second context sentence. Second, in the competitor-mentioned conditions (where the pronoun in the second sentence referred to the subject and may therefore have been strongly preferred to a noun phrase), there was no evidence that speakers took into account the addressee’s discourse model. If speakers took into account the referent’s accessibility in the addressee’s model, pronouns should have been less frequent when the addressee heard reference to the competitor than when only the speaker heard this. This is clearly not what we found; if anything, the direction of the results went in the opposite direction, consistent with our earlier suggestion that speakers may have engaged in what Galati and Brennan (2010) called “one-bit, most minimal audience design” rather than that they took accessibility for their addressee into account. Therefore, the lack of audience design cannot be attributed to the use of a pronoun in the target-mentioned conditions.

4. General discussion

Both Experiments 1 and 2 showed that the referent’s accessibility in the speaker’s own, privileged discourse model affected the choice of referring expressions. Experiment 1 found that in conditions where the preceding sentence was not shared with the addressee, speakers
nevertheless produced significantly more pronouns when this sentence mentioned the referent than when it mentioned the competitor, indicating that speakers used their own discourse model. Experiment 1 also showed a nonsignificant tendency for speakers to produce more pronouns when the addressee heard the sentence that mentioned the referent than when the addressee did not. However, Experiment 2 showed that this tendency occurred regardless of whether the preceding sentence mentioned the referent or the competitor. This suggests that speakers may have been somewhat more explicit when the preceding sentence was shared with their addressee than when it was not, but they did not take into account whether this sentence made the referent or the competitor accessible to the addressee. Most important, Experiment 2 showed that speakers produced more pronouns when the preceding sentence mentioned the referent compared to when it mentioned the competitor, and that this effect was equally large when the addressee could hear the preceding sentence as when the addressee could not hear it. This indicates that speakers chose the form of referring expression based on the referent’s accessibility in their own rather than the addressee’s discourse model; they did not take into account whether the second sentence made the referent more accessible to the addressee or only to themselves.

Our results thus provide evidence against the widely held assumption in linguistic theories of reference that speakers use the referent’s accessibility in the addressee’s discourse model when they choose referring expressions (e.g., Ariel, 1990; Chafe, 1994; Givón, 1983; Prince, 1985); there was no evidence that speakers choose referring expressions to signal how accessible a discourse entity is in the addressee’s discourse model. In contrast, our results support the view that the referent’s accessibility in the speaker’s own discourse model is the driving force behind the choice between pronouns and definite noun phrases. The results also have important implications for computational models designed to generate referring expressions produced by human speakers (Van Deemter, Gatt, Van Gompel, & Krahmer, 2012). As discussed earlier, computational theories of reference (e.g., Appelt, 1985a; Dale & Reiter, 1995) are often designed to model referring expressions that are optimally helpful to addressees, but they tend to pay little attention to the fact that speakers do not always choose referring expressions that are adapted for the addressee’s comprehension, as shown in our current study.

Our finding appears to contrast with studies which show evidence for audience design. However, most of these studies investigated cases where failing to take the addressee’s perspective would make reference ambiguous or that would result in referential failure. In studies showing that speakers take into account the addressee’s familiarity with a particular topic or referring expression (Brennan & Clark, 1996; Horton & Gerrig, 2002, 2005a; Isaacs & Clark, 1987; Wilkes-Gibbs & Clark, 1992), speakers had to adapt their reference to their addressee because they knew that the addressee, who was unfamiliar with the referent or the referring expression, might not be able to uniquely identify the referent. In studies showing that speakers use the addressee’s visual perspective to avoid referential ambiguity (Horton & Keysar, 1996; Nadig & Sedivy, 2002), disambiguation was necessary when the addressee could see the competitor, whereas it was not when the addressee could not see the competitor. Finally, in Matthews et al. (2006) study, which showed that 3- and 4-year-old children take the addressee’s visual perspective when using unheralded pronouns, unheralded
pronouns failed to unambiguously identify the referent when the referent was invisible to the addressee. In contrast, in our experiments, both the use of a definite noun phrase and a pronoun resulted in unambiguous reference, because the different gender of the referent and the competitor rendered pronouns unambiguous. Therefore, the use of a pronoun did not result in communicative failure that could have compelled the speaker to model the addressee’s representation.

One possibility is that, as suggested by many researchers, audience design depends upon the speaker’s awareness that audience design is needed. Ferreira, Slevc, and Rogers (2005) showed that speakers almost always produced modified noun phrases (e.g., *large bat*) when the context contained another same-category exemplar (e.g., *small flying bat*), whereas speakers often produced unmodified, ambiguous bare nouns (e.g., *bat*) in the context with a competitor that belonged to a different semantic category (e.g., baseball bat when the referent was a flying bat). Their results suggested that speakers can avoid ambiguity when ambiguity is obvious to speakers (e.g., when the context contains objects that are identical apart from size), but they often fail to avoid ambiguity when detecting ambiguity requires speakers to retrieve the linguistic forms for the referential candidates. Perhaps speakers are able to take the addressee’s visual perspective when the visual context contains a semantically identical referential competitor (Horton & Keysar, 1996; Nadig & Sedivy, 2002) because semantically based referential ambiguity may be obvious to speakers and they use the addressee’s visual perspective to avoid it. Similarly, during interactive dialogues, speakers may realize the needs of their addressee when they receive the addressee’s feedback in cases where the referring expression is uncommunicative (Horton & Gerrig, 2002) or when the addressee’s familiarity with the topic or the expression is strongly associated with the addressee’s identity (Horton & Gerrig, 2005a; Isaacs & Clark, 1987).

It is also possible that successful audience design depends upon the availability of the addressee’s knowledge state. In the studies that showed evidence of visual perspective-taking, the addressee’s visual perspective was not only clear from the configuration of the visual array, but it was also relevant for the speaker’s task: In standard referential communication tasks, privileged objects are never target referents because the addressee cannot manipulate them, so speakers may represent shared and privileged objects differently (Wardlow-Lane & Ferreira, 2008). Furthermore, Horton and Gerrig (2005a) argued that when selecting a conversational topic, speakers automatically retrieve information about events and individuals associated with the particular addressee, so speakers can take into account the addressee’s familiarity with a specific topic when choosing a referring expression. In contrast, keeping track of the accessibility of referents from the addressee’s perspective during a conversation is not so straightforward, because the linguistic context rapidly changes over the course of the discourse, so speakers need to keep in memory which sentence they heard with the addressee and what entity was mentioned in that sentence.

Thus, our results are consistent with accounts that assume that interlocutors use their “egocentric” model for spontaneous, routine linguistic processing, while the addressee’s knowledge and cognitive state have an influence only under limited circumstances (e.g., Barr & Keysar, 2006; Brown & Dell, 1987; Dell & Brown, 1991; Keysar, Lin, & Barr, 2003; Pickering & Garrod, 2004), when speakers are cued to realize that audience design is
necessary and they are very clear about their addressee’s cognitive state. Presumably, constant updates of the addressee’s current knowledge status are too cognitively demanding (Horton & Keysar, 1996; Rossnagel, 2000) to engage within the time frame of utterance planning (cf. Horton & Gerrig, 2005a). It is also possible that the choice between a pronoun and a repeated noun is made during what Bard and Aylett (2005) call fast automatic processes, whereas avoiding referential ambiguity may involve controlled processes, which tend to proceed more slowly and are therefore more likely to be affected by the speaker’s belief about the other’s knowledge.

As discussed earlier, traditional linguistic theories (e.g., Ariel, 1990; Chafe, 1994; Clark & Marshall, 1981; Givón, 1983; Gundel et al., 1993) often assume that speakers vary the form of reference to signal accessibility to the addressee. But if speakers are egocentric, why should they vary the form of reference at all? In our experiments, participants could have chosen pronouns as opposed to repeated nouns much more frequently, because pronouns are shorter, more frequent, and hence presumably much easier to produce than more explicit definite noun phrases (Almor, 1999; Ariel, 1990). In fact, if participants were egocentric, we might expect that they are more likely to produce definite noun phrases when their referent is more rather than less accessible. When the referent is highly accessible, it should be relatively easy to access a noun describing it, because the semantic information needed for lexical retrieval may be highly activated (e.g., it has just been mentioned) and its phonological form may also be easily accessible (repeating a recently mentioned word is usually easier than producing a word that has not been mentioned). In contrast, when the referent is less accessible, accessing a noun that describes it should be harder.

One possibility is that speakers sometimes avoid pronouns because they use their own discourse model as a proxy for their addressee’s (cf. Dell & Brown, 1991; Pickering & Garrod, 2004). When the referent is highly accessible in their own discourse model, speakers choose reduced referring expressions such as pronouns, because they assume that the referent is equally accessible in the addressee’s model and very little information is needed for identification, whereas when the referent is less accessible in their own discourse model, they choose more explicit referring expressions because they assume that the referent is less accessible in their addressee’s model and the addressee requires more information to identify the referent. In most cases, as discourse progresses, the interlocutors’ representations converge (cf., Brown & Dell, 1987; Pickering & Garrod, 2004), so what is accessible for the speaker is also accessible for the addressee. Because of this ‘alignment’ of speaking and listening, using one’s own perspective permits the production of referring expressions that are adapted for the addressee’s comprehension.

Although the above-mentioned account assumes that speakers use their own discourse model, it does postulate that speakers vary the form of reference to facilitate the addressee’s identification. However, it is also possible that accessibility effects result from a purely speaker-internal production process. Fukumura and Van Gompel (2011) argued that the choice of referring expression may be determined by how much conceptual encoding is needed when speakers initiate language production processes to refer to a discourse entity. According to models of language production (e.g., Bock & Levelt, 1994; Levelt, 1989), language production starts with the formulation of the conceptual representation a speaker
wants to express, which is assumed to guide subsequent linguistic encoding. Fukumura and
Van Gompel argued that when the referent is highly salient in the discourse context, the ref-
erent’s representation is already clearly represented in memory, so relatively little concep-
tual encoding is needed. As a result of this, speakers may be more likely to produce
referring expressions that contain little descriptive information about the referent such as
pronouns. Conversely, when the referent is less accessible, speakers may need to access
more conceptual features about the referent, which may result in the production of referring
expressions that contain more descriptive information about the referent such as definite
noun phrases. Fukumura et al. (2011) elaborated this view when they proposed a cue-based
retrieval model that assumes that the choice of referring expressions is determined by the
conceptual features speakers need to activate to identify the referent’s representation: The
less accessible or identifiable the referent is, the more conceptual features or attributes
speakers need to activate to identify the referent’s representation, which results in the
production of more elaborate referring expressions.

An alternative view that also explains the choice of referring expressions as a speaker-
internal factor claims that, similar to phonological or syntactic reductions, the choice of
referring expressions is determined by the word’s contextual probability (e.g., Aylett &
Turk, 2004, 2006; Jaeger, 2010; Levy & Jaeger, 2007). The basic idea is that the more prob-
able or predictable the word or the reference to the entity is in the context, the less informa-
tion the word conveys, so speakers tend to attenuate the phonological form of that word to
 avoid redundancy. Indeed, research has shown that the articulation of more predictable
words tends to be shorter and carries less phonetic detail (Aylett & Turk, 2004, 2006; Bell
et al., 2003; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999; Jurafsky, Bell,
Gregory, & Raymond, 2001). In this spirit, Jaeger (2010) argued that “speakers should be
more likely to produce pronouns (e.g., she) instead of full noun phrases (e.g., the girl) when
reference to the expression’s referent is probable in that context” (p. 48). However, it is
doubtful whether the choice of referring expressions is indeed affected by the likelihood of
reference. Fukumura and Van Gompel (2010) found that following Gary scared/feared
Anna because ..., participants used more pronouns (rather than names) to refer to Gary than
Anna, but their choice of referring expression was unaffected by the likelihood that they
referred to these discourse entities: They were more likely to refer to Gary than Anna when
the verb was scared, whereas they referred more often to Anna than Gary with feared, but
these verb biases had no effect on whether they used a pronoun or name. Therefore,
although research suggests that phonological and syntactic reduction are affected by predict-
ability, evidence suggests that the choice of referring expression is not.

In sum, we argue that speakers do not routinely use the addressee’s discourse model when
choosing between pronouns and definite noun phrases, but instead choose referring expres-
sions depending on how accessible the referent is in their own discourse model. The reason
for this may be that speakers use their own discourse model as a proxy for the addressee’s
or because the choice between pronouns and definite noun phrases is entirely driven by pro-
duction-internal constraints. Either way, given that during normal conversations, the inter-
locutors’ discourse representations often converge (cf. Brown & Dell, 1987; Pickering &
Garrod, 2004), using one’s own perspective generally suffices for producing referring
expressions that are adapted for the addressee’s comprehension. This obviates the need for constant, resource-demanding updates of the addressee’s perspective.

Acknowledgments

This study was supported by a University of Dundee Demonstratorship and Economic and Social Research Council Postdoctoral Fellowship PTA-026-27-2621 to the first author. We thank Emiel Krahmer for helpful suggestions.

References


**Appendix**

Context sentences used in Experiments 1 and 2, followed by the auditory sentences (the sentence before the slash represents the target-mentioned condition and the sentence after the slash the competitor-mentioned condition) and actions depicted in the photos (in uppercase).

The Viking is fishing in the lake with the nanny. She is lying down. He is lying down.

NANNY PUTTING BUCKET UPSIDE DOWN

The mermaid is waiting for a taxi with the admiral. He is sitting in a wheelchair. She is sitting on a bench.

ADMIRAL GETTING OFF WHEELCHAIR

The Indian is on the run with the lady. She is carrying a bag. He is carrying a bag.

LADY HOLDING CHAIN UP

The witch is taking a walk with the priest. He is holding an umbrella. She is holding a stick.

PRIEST TURNING AROUND

The wizard has cleaned the windows with the princess. She is holding a bucket. He is holding a bucket.

PRINCE SITTING DOWN

The girl is having fun with the king. He is standing on a chest. She is standing on a chest.

KING RAISING MEGAPHONE

The cowboy is approaching a castle with the policewoman. She is wearing a hat. He is wearing a hat.

POLICEWOMAN GETTING OFF HORSE

The gladiator is on the battleground with the countess. She is wearing a dress. He is wearing a helmet.

COUNTESS RAISING SWORD
The lady is discussing the weather with the policeman. He is standing next to a tree. She is standing next to a tree. POLICEMAN RAISING HANDS

The nanny is having a picnic with the prince. He is lying on the ground. She is lying on the ground. PRINCE STANDING UP

The maid has arrived at a park with the cowboy. He has got a gun. She has got a spoon. COWBOY PUTTING HAND INTO FOUNTAIN

The king is talking about politics with the nun. She is lying on the sofa. He is lying on the sofa. NUN BOWING

The policewoman is playing in a garden with the boy. He is wearing a blue shirt. She is wearing a white vest. BOY DROPPING ICE CREAM

The captain hurries to the station with the bride. She is holding a map. He is holding a map. BRIDE FALLING TO GROUND

The pilot is arguing about the traffic with the mermaid. She has blond hair. He has brown hair. MERMAID GETTING OUT OF CAR

The admiral is searching for treasure with the queen. She is wearing a crown. He is wearing a hat. QUEEN DROPPING CANDLE

The stewardess is visiting the village with the gladiator. He is looking back. She is looking back. GLADIATOR GETTING OFF HORSE

The boy is quarrelling about a road sign with the witch. She is raising a stick in the air. He is raising a hand in the air. WITCH GETTING INTO CAR

The countess is in the forest with the Indian. He has got a trumpet. She has got a trumpet. INDIAN SHOOTING

The sheriff is at the riverside with the girl. She is holding a balloon. He is holding a balloon. GIRL GETTING INTO CANOE

The queen is hiking in the hills with the gentleman. He has got a suitcase. She has got a camera. GENTLEMAN SITTING DOWN

The nun is pruning the roses with the wizard. He is wearing a gown. She is wearing a gown. WIZARD TURNING AROUND

The groom has practiced for a fire drill with the maid. She has got a fire-extinguisher. He has got a hose. MAID TURNING FIREEXTINGUISHER UPSIDE DOWN

The woman is in a pub with the pirate. He is wearing a yellow jacket. She is wearing a yellow jacket. PIRATE PUTTING UP HAND