

Comparing individual and collective management of referential choices in dialogue

Journal:	<i>Quarterly Journal of Experimental Psychology</i>
Manuscript ID	QJE-STD-20-402.R2
Manuscript Type:	Standard Article
Date Submitted by the Author:	12-Jul-2021
Complete List of Authors:	Knutsen, Dominique; Université de Lille, Fossard, Marion; Université de Neuchâtel, Institut des Sciences Logopédiques, Maison des Sciences du Langage et de la Communication Achim, Amélie; Université Laval, Département de psychiatrie et neurosciences
Keywords:	Dialogue, Referential choices, Grounding

SCHOLARONE™
Manuscripts

RUNNING HEAD: COLLECTIVE REFERENTIAL CHOICES IN DIALOGUE

Comparing individual and collective management of referential choices in dialogue

Dominique KNUTSEN

Univ. Lille, CNRS, CHU Lille, UMR 9193 - SCALab - Sciences Cognitives et Sciences

Affectives, F-59000 Lille, France

Marion FOSSARD

Institut des Sciences Logopédiques, Maison des Sciences du Langage et de la Communication,

Université de Neuchâtel, Switzerland

Amélie M. ACHIM

Département de psychiatrie et neurosciences, Université Laval, Québec, Canada

Centre de recherche CERVO, Québec, Canada

Corresponding author:

Dr Dominique Knutsen

Laboratoire SCALab UMR CNRS 9193

Université de Lille

Rue du barreau

BP 60149, 59653 Villeneuve d'Ascq Cedex

France

Email address: dominique.knutsen@univ-lille.fr

Abstract

Past research shows that when a discourse referent is mentioned repeatedly, it is usually introduced with a full NP and maintained with a reduced form such as a pronoun. Is this also the case in dialogue, where the same referent may be introduced by one person and maintained by another person? An experiment was conducted in which participants either told entire stories to each other or told stories together, thus enabling us to contrast situations in which characters were introduced and maintained by the same person (control condition) and situations in which the introduction and the maintaining of each character were performed by different people (alternating condition). Story complexity was also manipulated through the introduction of one or two characters in each story. We found that participants were less likely to use reduced forms to maintain referents in the alternating condition. The use of reduced forms also depended on the context in which the referent was maintained (in particular, first or second mention of a character) and on story complexity. These results shed light on how the pressure to signal understanding to one's conversational partner affects referential choices throughout the interaction.

Keywords

Dialogue; referential choices; referential accessibility; grounding; turn-taking

Highlights

- This experiment examined referential choices in interactive dialogue.
- Participants took turns introducing and maintaining characters in stories.
- The opportunity to take turns led dialogue partners to produce fewer reduced forms.
- The pressure to ground information affects referential choices in dialogue.

Comparing individual and collective management of referential choices in dialogue

1. Introduction

There are usually many different ways of referring to the same thing. For instance, the same person may be referred to as “the tall woman”, “the lady”, “the math teacher” or “she”, implying that speakers must make a number of referential choices before producing a referential expression (Gatt et al., 2014). Such choices concern the underlying conceptualization of the referent (e.g., the “lady” versus the “math teacher”), the content of the referential expression (i.e., the extent of the information provided in the noun phrase; e.g., the “tall woman” versus the “woman”) as well as the reference marker used (i.e., whether the referential expression is definite, indefinite, or is a pronoun; e.g., “the tall woman”, “a tall woman” or “she”).

Over the past decades, different theoretical models (Ariel, 1990, 2001; Givon, 1983; Gundel et al., 1993) have sought to identify the factors which affect such referential choices in discourse. With a specific focus on the choice of reference markers used, studies in this field - in particular in narrative contexts - have shown that referent accessibility plays a crucial role. Typically, when a referent is new or not already accessible in the discourse representation, a full and explicit noun phrase (NP), either indefinite or definite, is generally used. But when a referent is readily accessible in the discourse representation, the speaker is more likely to choose a full NP with no modifier (e.g., “the woman”) or a reduced form such as a pronoun (e.g., “she”) or even a zero anaphora (Arnold, 2010; Arnold & Griffin, 2007; Hendriks et al., 2014). Referents that are given (i.e., that have already been mentioned earlier in the discourse) and topical (i.e., that have been mentioned *recently* in the discourse, in a syntactically prominent position) tend to be particularly accessible in the discourse model (Grosz et al., 1995) and are thus usually referred to using a reduced form. For instance, Colle et al. (2008) used a storytelling procedure in which participants were asked to look at the pictures from a 24 picture booklet, and to turn the pages whilst telling the story to an experimenter (who did not intervene during the narration). In that study, participants typically used an indefinite full NP when introducing a character (a boy) for the first time in the discourse; and pronouns were then used to refer to him again as long as he remained in the participants’ focus of attention; finally, full NPs (indefinite + definite NPs) were used to reintroduce this character after a different character had been referred to in the discourse. In a similar way, Hendriks et al. (2014) examined narratives which featured two characters. In that study,

1
2
3 participants were asked to tell stories based on 6 pictures storybooks to a hypothetical listener. The
4 first character was usually introduced using a full NP and maintained (i.e., referred to again) using a
5 pronoun. Then, the second character was both introduced *and* maintained using a full NP – and not a
6 pronoun as expected (Hendriks et al. suggested that the second character had not yet been identified
7 as topical in the discourse, explaining why the participants did not switch to a reduced form to maintain
8 the referent). Finally, the first character was introduced again, or reintroduced, using a full NP again.

14 Another situation in which speakers must make referential choices is dialogue, in which
15 speakers have the opportunity to jointly refer to things such as objects, entities, people or animals (see
16 for instance Brennan & Clark, 1996; Clark & Bangerter, 2004; Clark & Schaefer, 1989; Clark & Wilkes-
17 Gibbs, 1986; Goudbeek & Kraemer, 2012; Horton & Gerrig, 2002; Ibarra & Tanenhaus, 2016; Knutsen
18 et al., 2018; Viethen et al., 2014; Yoon & Brown-Schmidt, 2014). Although a number of dialogue
19 studies have focused on referential choices at the conceptual level (e.g., whether an abstract picture
20 such as a tangram figure should be referred to as a person or an animal; Brennan & Clark, 1996;
21 Clark & Wilkes-Gibbs, 1986), these studies tended to examine referential expression content (i.e., the
22 choice of properties or attributes <colour, size> that are added to the expression in order to foster
23 referent identification; Koolen et al., 2011; Viethen et al., 2014) or to examine whether dialogue
24 partners use definite or indefinite referential markers as they interact (see for instance Clark &
25 Marshall, 1981; Clark & Wilkes-Gibbs, 1986). In contrast, only a small number of studies have focused
26 on the use of reduced forms during dialogue. One potential explanation is that most dialogue studies
27 use identification tasks in order to examine the processes underlying reference production in dialogue.
28 In these tasks, a participant must produce a referential expression in order to enable his or her partner
29 to find the corresponding referent (e.g., a picture or an object) in a set of potential referents (e.g.,
30 Achim et al., 2015; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986). Such settings indirectly
31 encourage the use of full forms (e.g., definite or indefinite full NPs, rather than reduced forms such as
32 pronouns) because each trial targets a new referent and participants need specific information about
33 the features of that referent in order to be able to distinguish it from potential competitors.

52 To sum up, the studies that allowed for dialogue between participants tend to use identification
53 tasks, and the studies focusing on discourse and covering a wider range of reference markers usually
54 have participants tell their stories to listeners who do not intervene during the narration. This highlights
55 the need to develop new methodologies that address the interactive and continuous nature of dialogue
56
57
58
59
60

1
2
3 and that also covers the full range of discourse stages (introduction, maintaining and reintroduction or
4 shift), enabling researchers to examine participants' use of reduced forms (or not) that appear
5 principally during the maintaining stage.
6
7

8
9 A first step towards this comes from the study by Fossard et al. (2018), who examined the
10 referential choices (referential expression content and markers) made by French-speaking participants
11 during a new storytelling task that makes a clearer distinction between the three discourse stages.
12
13 Importantly, this storytelling task uses a referential communication paradigm in which two dialogue
14 partners must collaborate to reach a common goal (Clark & Wilkes-Gibbs, 1986, see also Achim et al.,
15 2017). While this paradigm has been widely used in previous target identification task, to our
16 knowledge only one study had previously used the referential communication paradigm in the context
17 of a storytelling task (Achim et al. 2017).
18
19
20
21
22
23

24 In both Achim et al. (2017) and Fossard et al. (2018), the participant received sequences of six
25 images in a predetermined order and was instructed to tell the corresponding story to a research
26 assistant playing the role of the addressee. The latter, who was separated from the participant by an
27 opaque screen, held the same set of images in a random order and had to place them in the correct
28 order based on the story told by the participant. In contrast to the storytelling studies presented above
29 (Colle et al., 2008; Hendriks et al., 2014), the studies by Achim et al (2017) and Fossard et al. (2018)
30 were genuinely collaborative, as the addressee had to use the information provided by the speaker in
31 order to place the images in the same order as the participant-narrator. What is more, during the task,
32 the addressee could provide feedback not only to signal understanding (e.g., "mhm", "okay") but also
33 to ask for clarifications in case there was an ambiguity (e.g., "can you give me more detail?").
34
35
36
37
38
39
40
41
42

43 The study by Fossard et al. also manipulated the number of characters in the story (one or
44 two), and when there were two characters, they manipulated their visual salience and level of activity
45 (which impact the choice of referent) and their gender (either of the same gender or of opposite
46 genders) in order to manipulate referential complexity. Single character stories featured one character
47 only, gender unambiguous stories featured two characters of opposite genders and gender ambiguous
48 stories featured two characters of the same gender. In single character stories, the character was
49 introduced in picture 1 and maintained in focus in pictures 2-6. In gender unambiguous and gender
50 ambiguous stories, pictures 1 and 2 focused on one character (hereafter referred to as character 1,
51 who was the visually salient, agent character in these pictures), pictures 3 and 4 focused on the
52
53
54
55
56
57
58
59
60

1
2
3 second character (hereafter referred to as character 2, who became the visually salient character in
4 these pictures) and pictures 5 and 6 focused on character 1 again. An example is provided in Figure 1
5
6 below.
7
8
9

10
11 --- Insert Figure 1 about here ---
12
13
14
15

16 The authors examined the referential expressions used by the participants to refer to the
17 character in focus. Specifically, they looked at the referential expressions at different discourse stages,
18 namely the expressions used to introduce the character in picture 1 (introduction stage), to shift to a
19 different character in pictures 3 and 5 (only for gender unambiguous and gender ambiguous stories;
20 shift stage) and to maintain a character in pictures 2-6 for single character stories, and in pictures 2, 4
21 and 6 for gender unambiguous and ambiguous stories (maintaining stage). The results revealed that
22 both discourse stages and referential complexity affected which reference markers (i.e., indefinite NP,
23 definite NP, or pronoun) were produced by the participants. In a nutshell, indefinite NPs were mainly
24 used in the introduction stage and definite NPs were mainly used in the shift stages (which applies
25 only for gender unambiguous and gender ambiguous stories). As for reduced forms, they were mainly
26 used in the maintaining stages.
27
28
29
30
31
32
33
34
35
36

37 It is noteworthy that although Fossard et al.'s (2018) study involved genuine interactions
38 between participants and research assistants, referential choices were managed individually in that
39 study, that is, all critical referential expressions to the characters in the stories were systematically
40 produced by the same person (i.e., the participant). This might not always be the case in spontaneous
41 dialogue, where the management of referential choices may be *collective*. In the current paper, we
42 define situations in which referential choices are managed collectively as situations in which one of the
43 speakers introduces a referent, which is then maintained by a different speaker. In such cases,
44 although referential decisions are *in fine* made by only one person (namely, the current speaker), the
45 important point is that both partners contribute actively to the chain of referential expressions.
46
47
48
49
50
51
52
53
54

55 Such collective management of referential decisions could have important implications for the
56 study of referential choices made by dialogue partners, and more specifically for the choice to use a
57 full NP versus a reduced form such as a pronoun or a zero anaphora in the maintaining stage. Indeed,
58
59
60

1
2
3 when two people interact, they must provide cues showing that they believe that they have understood
4 each other correctly throughout the entire interaction (Clark & Brennan, 1991; Clark & Schaefer,
5 1989). Doing so enables them to *ground* the information under discussion, that is, to add this
6 information to their common ground, which includes the knowledge which they are aware of sharing
7 (Clark & Brennan, 1991; Clark & Marshall, 1981; Fox Tree & Clark, 2013; Gergle et al., 2013;
8 Lysander & Horton, 2012). Depending on how much knowledge is shared by both partners,
9 information may be grounded immediately after its presentation (if the addressee believes to have
10 understood what was said well enough for current purposes) or several speech turns later (if the
11 addressee requires more information before he or she can confirm that he or she understood what
12 was said) (Clark & Brennan, 1991; Clark & Schaefer, 1989). What is more, grounding is more or less
13 explicit: an addressee may signal his or her understanding of what was said by repeating all or part of
14 their partner's utterance, by saying "yes", "no" or "okay", by nodding or smiling, or even by initiating the
15 next relevant speech turn (Clark & Schaefer, 1989; Knutsen & Le Bigot, 2012, 2014, 2015).
16 Importantly, research has shown that dialogue partners tend to favor the use of explicit feedback when
17 they feel that the risk of misunderstanding each other is high (Knutsen & Le Bigot, 2012; McInnes &
18 Attwater, 2004; Roque & Traum, 2008). For instance, when two people exchange phone numbers,
19 they tend to spontaneously repeat the digits out loud in order to signal their understanding (and to
20 provide the other person the opportunity to correct them if a digit was misunderstood at some point).
21 This has important implications for situations in which a referent is introduced by one person and
22 maintained by someone else. Indeed, also producing a full NP¹ rather than switching to a pronoun or a
23 zero anaphora to maintain a referent that was introduced by someone else might help ensure mutual
24 comprehension. For instance, if Speaker A produces the full NP "the teacher" during a dialogue and
25 Speaker B then maintains this reference by producing the same NP, this might be because Speaker B
26 wants to show Speaker A that she has understood the reference correctly. In other words, in dialogue,
27 a speaker might *avoid* using a pronoun or a zero anaphora when maintaining a referent that was
28 initially introduced by someone else as a strategy to reach mutual understanding, suggesting that

29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
¹ Although the full NP used to maintain the referent would obviously be used to refer to the same *referent*, the *referential expression* used to maintain this referent might be a different one to the one initially produced. For instance, a speaker might maintain a referent referred to as "the little boy" by saying "the little boy" or "this guy".

1
2
3 different strategies may be used by speakers to maintain referents depending on whether referential
4
5 choice management is collective or individual.
6
7

8 9 **1.1. Overview of the experiment and hypotheses**

10 The purpose of the current work is to examine how people maintain referents in dialogic situations
11 where participants take turns producing referential expressions (collective management of referential
12 choices), and to compare these situations to dialogic situations where all referential expressions are
13 produced by the same person (individual management of referential choices). In order to do this, we
14 used an adapted version of the collaborative storytelling task developed by Fossard et al. (2018). In
15 the *alternating* condition, pairs of participants told the story jointly. In single character stories, this
16 meant that the referential expression to the character in focus in the introduction was generated by
17 one of the participants; both participants then took turns to maintain this character. In gender
18 unambiguous and gender ambiguous stories (i.e., stories which featured two characters), this meant
19 that one of the participants produced referential expressions to introduce character 1 in picture 1, to
20 shift to character 2 in picture 3 and to shift back to character 1 in picture 5; as for the other participant,
21 he or she maintained character 1 in picture 2, character 2 in picture 4 and character 1 again in picture
22 6. In the *control* condition, the entire story was told by one participant to another participant, such that
23 all referential expressions to the characters in the stories were produced by the same person (see
24 Figures 1-3).
25
26
27
28
29
30
31
32
33
34
35
36
37
38

39 The dialogues between the participants were examined to determine whether referents were
40 maintained (i.e., during the maintaining stages) by using reduced forms (pronoun or zero anaphora)
41 rather than full NPs. The main hypothesis was that in the alternating condition, participants use fewer
42 reduced forms to maintain referents than in the control condition. This pattern would be expected if
43 participants prioritize the production of full NPs as a grounding strategy in the alternating condition.
44
45
46
47
48

49 Moreover, the collaborative storytelling task enabled us to characterize this grounding process
50 across different maintaining contexts and across different levels of referential complexity. Specifically,
51 examining the maintaining of referential expressions after the introduction of character 1, after the shift
52 to character 2 and after the shift back to character 1 enabled us to determine whether grounding
53 through the production of a full NP occurs mainly at the beginning of the dialogue, when the
54 characters in the story have not been added to the participants' common ground yet, or whether
55
56
57
58
59
60

1
2
3 participants tend to ground information through the production of a full NP throughout the entire
4
5 interaction. As for referential complexity, we sought to determine whether grounding through the
6
7 production of a full NP occurred mainly when the level of complexity is high, which could indicate that
8
9 dialogue partners become more likely to repeat information when the risk of miscomprehension
10
11 increases (see McInnes & Attwater, 2004).

14 2. Method

16 2.1. Participants

17
18 Fifty participants (mean age 19.46 years, $SD = 1.96$, 6 men) divided into 25 pairs were tested in this
19
20 experiment. All were students at the University of Essex (UK). Although some were bilingual or
21
22 multilingual, they all confirmed that English was their native language. Pairs of participants did not
23
24 know each other prior to the experiment. They received either course credits or a small payment (£5)
25
26 in exchange for their participation. They signed an informed consent form at the beginning of the
27
28 experiment and were fully debriefed afterwards.

31 2.2. Materials

32
33 The materials used were adapted from Fossard et al. (2018). A total of 12 stories were used, four of
34
35 which featured either one male or one female character (single character stories) and eight of which
36
37 featured two characters (one male and one female in four stories [gender unambiguous stories] and
38
39 two males or two females in the other four stories [gender ambiguous stories]). Each story was made
40
41 up of six pictures, as shown in Figure 1.

42
43 For the intermediate and complex stories, character 1 was made salient in picture 1
44
45 (introduction stage) by being depicted performing an action on the picture and remained salient (in the
46
47 foreground and active) in picture 2 (first maintaining of character 1); character 2 was only visible in the
48
49 background of picture 2. Character 2 was then made salient in picture 3 (shift to character 2) and
50
51 picture 4 (first maintaining of character 2), whereas character 1 was shown in the background of both
52
53 pictures. Finally, character 1 was made salient again in picture 5 (shift back to character 1) and picture
54
55 6 (second maintaining of character 1). This manipulation has proved successful at encouraging
56
57 participants to refer to character 1 in pictures 1, 2, 5 and 6, and to refer to character 2 in pictures 3 and
58
59 4. In contrast, simple stories only involved one character. Thus, in simple stories, picture 1
60

1
2
3 corresponded to the introduction stage whereas all other pictures corresponded to the maintaining of
4
5 character 1.

6
7 Different versions of each story were then created for use in both conditions (i.e. the materials
8
9 were counterbalanced so that each story was used in the control condition and the alternating
10
11 condition the same number of times across all pairs). In the control condition, the participant telling the
12
13 story (hereafter referred to as the director) received the entire story, printed on an A4 sheet of paper.
14
15 The other participant (hereafter referred to as the matcher) received an empty 3 x 2 grid, which was
16
17 also printed on an A4 sheet of paper, as well as the same six pictures as the ones shown to the
18
19 director, but these were printed on loose cards, as shown in Figure 2.

20
21
22 --- Insert Figure 2 about here ---
23
24
25

26
27 In the alternating condition, one of the participants received a grid including pictures number 1,
28
29 3 and 5, as well as three loose cards representing pictures 2, 4 and 6. The other participant received a
30
31 grid including pictures 2, 4 and 6 of the same story (i.e. all the pictures for the maintaining stage), as
32
33 well as three loose cards representing pictures 1, 3 and 5. An example of these grids is provided in
34
35 Figure 3.

36
37
38 --- Insert Figure 3 about here ---
39
40
41
42

43
44 The dialogues between the participants were recorded using a double entry Tascam DR40
45
46 digital voice recorder.
47
48

49 **2.3. Procedure and method**

50
51 The experiment took place in a quiet room. The two participants sat at different desks facing different
52
53 sides of the room, so that they could not see each other as they performed the task. After they both
54
55 signed an informed consent form, the participants took part in a storytelling game.

56
57 For the trials corresponding to the control condition, the director's task was to tell the story so
58
59 that the matcher could place his or her cards in the right order in the empty grid. In other words, the
60

1
2
3 procedure in this condition was the same as in the initial study by Fossard et al. (2018), except that
4 the role of the addressee was held by a real naïve participant instead of a research assistant.
5

6
7 For the trials of the alternating condition, the participants took turns at playing the roles of
8 director and matcher within a same story. Specifically, the participants were told that the participant
9 who saw the first picture in his or her grid was the director for the first picture, and had to start the
10 story so that the matcher could find the first picture and place it in his or her own grid. Once this was
11 done, the other participant, who saw the second picture in his or her grid, became the director for this
12 picture, and his or her partner became the matcher. The participants switched roles this way until the
13 entire story had been told.
14
15
16
17
18
19

20 In both conditions, the experimenter emphasized that the participants' task was to tell a story,
21 rather than describe the pictures one by one. For this reason, the participants were asked not to
22 mention the number of the images ("first image", "image two", etc.). What is more, the participant
23 playing the role of the matcher could produce as much feedback as he or she wished, and could ask
24 the director any clarification questions he or she felt were necessary. Once the participants had
25 finished telling the story, the director and the matcher were allowed to look at each other's grids and
26 cards to check that they had successfully completed the task. They then embarked on the next trial.
27
28
29
30
31
32
33

34 The participants performed 12 trials in total, each corresponding to a different story. Thus,
35 each dyad was exposed to six stories per condition, including two trials for each of the three levels of
36 complexity. The participants switched to a different condition every three trials, with the condition order
37 and the stories belonging to each block both randomized across dyads. At the beginning of each set of
38 three trials, the experimenter announced whether this was a set of trials where only one participant
39 would tell the story (in which case she also announced who this participant would be), or whether this
40 was a set of trials where both participants would tell the story together (in which case she also
41 announced which participant would start telling the story). In the control condition, each participant
42 played the role of the director three times (i.e. for one block) and the role of the matcher three times
43 (i.e. for the other block). In the alternating condition, each participant was the director for pictures 1, 3
44 and 5 three times (i.e., for one block) and for pictures 2, 4 and 6 three times (i.e., for the other block).
45
46
47
48
49
50
51
52
53
54

55 The experiment lasted between half an hour and 45 minutes. The participants were fully
56 debriefed before they left the room.
57
58
59
60

2.4. Data coding and dependent variable

Performance at the task was not examined, as the participants almost never committed any mistakes (e.g., misplacing a card) in the current experiment.

The utterances produced by the designated directors were transcribed. Matcher speech mainly involved backchannel feedback (e.g., “yeah”, “okay” or “got you”) and was not usually transcribed, except if the matcher made a comment or asked the director a question, as this experiment included no hypothesis about the utterances produced by the matcher. Dialogue samples are provided in Tables 1 (control condition) and 2 (alternating condition).

--- Insert Table 1 about here ---

--- Insert Table 2 about here ---

Referential expressions used to refer to the salient character for the maintaining stages (i.e., pictures 2-6 in single character stories and pictures 2, 4 and 6 in gender unambiguous and gender ambiguous stories) were identified and coded as to whether they were reduced forms (pronouns or zero anaphora), or not². Following Fossard et al.'s procedure (2018), referential expressions produced to refer to the non-salient character were not taken into account in the data analysis. Plural referential expressions (e.g., “the two guys” or “they”) were also discarded from further analysis. If a participant produced several referential expressions to refer to the salient character while describing the same picture, only the referential expression which included more specific information (this was typically the first expression produced) was taken into account in the analysis (e.g. ‘the shop assistant’ in the following example: “the shop assistant is carrying a lot of boxes it looks like she’s struggling a bit”). Coding examples are provided in Tables 3 (control condition) and 4 (alternating condition). The

² Whereas the current study only focused on the references produced during the maintaining stage, we plan on using the data from all three stages (introduction, maintaining and shift stages) in future work on referential choices in dialogue. In this context, each reference produced during the experiment (in all three stages) was also coded as to whether it was a full indefinite NP (e.g., “a (young) boy”), a full definite NP (e.g., “the (tall) woman”), an unaccented pronoun (“he”) or a zero anaphora (“and in revenge \emptyset throws sand”). Full definite NPs were then coded as to whether they began with a possessive marker (“his friend”), a demonstrative marker (“this person”) or an accented pronoun (e.g., HE). Full NPs were also coded as to whether or not they included a modifier of the head noun such as an adjective (e.g., the blond girl) or a prepositional phrase (e.g., the girl with blond hairs), for instance. For the purpose of the current analyses, these categories were then used to determine whether or not a pronoun or zero anaphora was used to refer to the salient character. Although the reference marker data are not presented in more detail in the current paper, as our hypothesis focused exclusively on referent maintaining, the entire dataset is available upon request.

1
2
3 reduced form coding shown in these tables was used as the binary dependent variable (DV) in the
4 analyses reported below. A preliminary inspection of the data revealed that when participants used a
5 reduced form, they used a pronoun in a large majority of cases (96.24% of cases; zero anaphora were
6 only used in 3.76% of cases).
7
8
9

10 The data from three pairs were double coded by two research assistants. The inter-rater
11 agreement (calculated using Cohen's Kappa) was $\kappa = 0.92$, reflecting almost perfect agreement. All
12 disagreements were solved through discussion and the remainder of the data were single-coded by
13 one of the research assistants.
14
15
16
17

18
19
20 --- Insert Table 3 about here ---
21

22 --- Insert Table 4 about here ---
23
24
25
26
27

28 **2.5. Experimental design and data analysis**

29
30 The first independent variable (IV) used in this experiment was the condition according to which each
31 story was told (alternating or control). The second IV was the level of complexity of each story (simple:
32 single character; intermediary: gender unambiguous; complex: gender ambiguous). The third IV was
33 the maintaining context (first maintaining of character 1, first maintaining of character 2, second
34 maintaining of character 1). All three IVs were within-dyads and within-participants.
35
36
37
38

39 All analyses reported below were conducted in SAS 9.3. The main hypothesis (on the effect of
40 speech turn-taking on referential choices) was tested in an analysis which included the data from all
41 three levels of referential complexity. In this analysis, the IVs were the condition and referential
42 complexity, and the DV was the probability of producing a reduced form to maintain a referent. What is
43 more, the effect of the maintaining context on referential choices was examined in a second analysis
44 that only included the data from gender unambiguous and gender ambiguous stories. Indeed, as
45 mentioned previously, single character stories did not include a shift to character 2 stage or a shift
46 back to character 1 stage and could therefore not be used to test the effect of the different maintaining
47 contexts. This analysis was similar to analysis 1, but it included an additional IV, maintaining context.
48
49 The data were analyzed using logistic mixed models. Mixed models were used to account for the fact
50 that participants were nested within dyads in this experiment. As for logistic models, they were used
51
52
53
54
55
56
57
58
59
60

1
2
3 because the DVs were binary (a reduced form was produced during the maintaining stages, or not).
4
5 Because logistic models were used, we report odds ratios (*OR*) rather than effect sizes. It is
6
7 noteworthy that *ORs* provide information similar to that included in effect sizes. For instance, an *OR* of
8
9 2.00 in favor of the control condition would imply that participants are twice as likely to use a reduced
10
11 form in the control condition than in the alternating condition. Confidence intervals (*CI*) corresponding
12
13 to the odd ratios are also reported below.

14
15 According to Barr, Levy, Scheepers, and Tily (2013), mixed models should include the
16
17 maximal random effects structure justified by the experimental design. Random effects enabled us to
18
19 account for the fact that analysis units such as participants, items (i.e., stories) and, in the current
20
21 experiment, dyads, differ from each other, and also differ in their sensitivity to within-dyad, within-
22
23 participant and/or within-item IVs in the design. However, statistical models which include the maximal
24
25 random effects structure often fail to converge. In order to overcome this issue in the current analysis,
26
27 we identified all the random effects which caused convergence issues (this is done automatically in
28
29 SAS; Kiernan, Tao, & Gibbs, 2012), we removed them, and then we ran the analysis again. The
30
31 results reported below thus correspond to the latter analysis, that is, the analysis in which random
32
33 slopes and intercepts causing convergence issues had been removed. The final random effects
34
35 structure used in each model is specified below.

36
37 The data from eight observations were removed from the analysis due to the participants not
38
39 mentioning the character in focus or using a plural reference (which could either imply that they were
40
41 referring to both characters or using “they” as a neutral pronoun), or to recording issues. Because of
42
43 this, the number of observations was not the same for all cells of the design (in analysis 1, the
44
45 imbalance was also caused by the fact that there were five maintaining images in single character
46
47 stories, but only three in gender unambiguous and gender ambiguous stories). The Satterthwaite
48
49 correction was thus used to estimate the degrees of freedom in the analyses. Importantly, both this
50
51 correction and the fact that some models included random slopes explains why the degrees of
52
53 freedom reported are different from those that would have been obtained in an ANOVA or a
54
55 regression.
56
57
58
59
60

3. Results

3.1. Analysis 1: Effect of condition and referential complexity on the production of reduced forms

The first analysis included condition, complexity and the interaction between these two factors as fixed effects. The binary outcome variable was whether a reduced form was produced, or not. The results (random effects structure used as well as coefficients associated with the random effects and fixed effects) are shown in Tables 5 and 6. The data are shown in Figure 4. As presented in more details below, this figure illustrates that participants were more likely to produce reduced forms to maintain referents in the control condition than in the alternating condition, and that reduced forms were more likely to be produced in single character stories than in gender unambiguous stories, and in gender unambiguous stories than in gender ambiguous stories.

--- Insert Table 5 about here ---

--- Insert Table 6 about here ---

--- Insert Figure 4 about here ---

A significant main effect of condition was found, $F(1, 27) = 28.17, p < .001$. Reduced forms were less likely to be produced in the alternating condition than in the control condition, $OR = 0.25, CI_{.95} = 0.15, 0.43$, in line with our hypothesis.

A significant main effect of referential complexity was also found, $F(2, 9) = 29.01, p < .001$. Reduced forms were less likely to be produced in gender unambiguous and gender ambiguous stories than in single character stories, respectively $OR = 0.16, CI_{.95} = 0.07, 0.37$ and $OR = 0.06, CI_{.95} = 0.03, 0.13$. An additional pairwise comparison (conducted using the LSMEANS statement in SAS, which is based on a t -test; Bonferroni-corrected) revealed no significant difference between gender unambiguous and gender ambiguous stories, $t(7) = -3.03, corrected p = .054$. This pattern replicates Fossard et al.'s (2018) finding that fewer reduced forms are produced when a narrative features two characters.

Finally, the condition x complexity interaction did not reach statistical significance, $F(2, 1092) = 2.39, p = .093$.

3.2. Analysis 2: Effect of condition, referential complexity and maintaining context on the production of reduced forms in stories with two characters

The second analysis included condition, complexity, maintaining context and all interactions between these factors as fixed effects. As in the first analysis, the binary outcome variable was whether a reduced form was produced, or not. The results (random effects structure used as well as coefficients associated with the random effects and fixed effects) are shown in Tables 7 and 8. The data are shown in Figure 5. Once again, this figure suggests that participants were more likely to produce reduced forms to maintain referents in the control condition than in the alternating condition. In gender unambiguous stories, reduced forms were more likely to be produced to maintain character 1 (either for the first time or the second time) than to maintain character 2. The same pattern was visible in gender ambiguous stories – in fact, the effect of maintaining context was potentially stronger in this condition, in which the likelihood of producing a reduced form to maintain character 2 was particularly low (i.e., below 0.30).

--- Insert Table 7 about here ---

--- Insert Table 8 about here ---

--- Insert Figure 5 about here ---

A significant main effect of condition was found, $F(1, 89) = 23.38, p < .001$. Reduced forms were less likely to be produced in the alternating condition than in the control condition, $OR = 0.24, CI_{.95} = 0.13, 0.43$.

A significant main effect of referential complexity was also found, $F(1, 7) = 9.25, p = .020$. Reduced forms were more likely to be produced in gender unambiguous stories than in gender ambiguous stories, $OR = 0.27, CI_{.95} = 0.10, 0.76$. The difference between the first analysis (where the difference between gender unambiguous and gender ambiguous stories did not reach statistical significance) and the second analysis is due to the fact that whereas the first analysis focused on the

entire dataset, the second analysis focused on gender unambiguous and gender ambiguous stories only and distinguished different maintaining contexts.

A significant main effect of maintaining context was also found, $F(2, 17) = 10.88, p = .001$. Reduced forms were less likely to be produced during the first maintaining of character 2 than during the second maintaining of character 1, $OR = 0.27, CI_{.95} = 0.10, 0.70$. There was no significant difference between the first maintaining of character 1 and the second maintaining of character 1, $OR = 2.62, CI_{.95} = 0.92, 7.45$. An additional Bonferroni-corrected pairwise comparison revealed a significant difference between the first maintaining of character 1 and the first maintaining of character 2, $t(20) = 4.53, corrected p = .001$.

Finally, there was a significant interaction between condition x maintaining context, $F(2, 587) = 3.22, p = .041$. Pairwise comparisons (Bonferroni-corrected) revealed no significant effect of maintaining context in the alternating condition (i.e., there was no significant difference between the first maintaining of character 1 and the first maintaining of character 2, $t(20) = 2.87, adjusted p = .141$, between the first maintaining of character 2 and the second maintaining of character 1, $t(20) = -1.83, adjusted p < 0.99$, or between the first maintaining of character 1 and the second maintaining of character 1, $t(19) = 1.05, p < 0.99$). In contrast, in the control condition, reduced forms were significantly more likely to be produced during the first maintaining of character 1 than during the first maintaining of character 2, $t(82) = 5.18, adjusted p < .001$; reduced forms were also more likely to be produced during the second maintaining of character 1 than during the first maintaining of character 2, $t(22) = -3.33, adjusted p = .045$; the difference between the first maintaining of character 1 and the second maintaining of character 1 was non-significant, $t(81) = 1.99, adjusted p = .753$. Moreover, the difference between the alternating condition and the control condition was significant for the first maintaining of character 1, ($t(587) = -3.73, corrected p = .003$) and for the second maintaining of character 1 ($t(257) = -3.59, corrected p = .006$), but not for the first maintaining of character 2 ($t(216) = -1.64, corrected p = 1.00$).

All other effects failed to reach statistical significance, all p values $> .05^3$.

³ Specifically, the complexity x condition interaction was non-significant, $F(1, 587) = 3.15, p = .077$; the maintaining context x complexity was non-significant, $F(2, 16) = 1.56, p = .239$ and the condition x complexity x maintaining context was non-significant, $F(2, 587) = 2.79, p = .062$.

4. Discussion

The purpose of the current study was to examine the use of reduced referential forms in dialogic settings in which referential choice management is collective. Indeed, due to turn-taking, a referent may be introduced by one speaker and maintained by another speaker during dialogue. The hypothesis was that speakers are less likely to use reduced forms such as pronouns or zero anaphora (and thus using more full NPs) when maintaining a referent which was initially introduced by someone else, thus improving mutual comprehension (see Brennan & Clark, 1996; Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986). Indeed, also producing a full NP verbatim rather than switching to a reduced form may be used as a strategy to ground information in dialogue, that is, to signal to the other person that the referential expression was understood correctly (e.g., Clark & Brennan, 1991; Knutsen & Le Bigot, 2012; McInnes & Attwater, 2004; Roque & Traum, 2008). The findings reported here are in line with this hypothesis, as the participants in this study produced fewer reduced forms (and thus more full NPs) in the alternating condition than in the control condition. This highlights how referential choices may differ depending on the features of the dialogue situation in which a referent is maintained (i.e., in a collective or individual way). When participants produced reduced forms, they used a pronoun in most cases.

It is important to highlight that although our initial hypothesis was that dialogue partners would avoid using reduced forms (and hence use more full NPs) to maintain referents when referential choice management is collective, the current study does not enable us to determine whether such behavior is intentional or not. Grounding is usually defined as a strategic, costly process (e.g., Horton & Keysar, 1996; Rosnagel, 2000). However, the current findings might be attributable to a less costly, automatic priming process, whereby words mentioned earlier in the dialogue become more likely to be repeated due to ease of access in memory (e.g., Pickering & Garrod, 2004). Regardless of whether production of full NPs was strategic or automatic in the current study, its impact on mutual comprehension would have been the same, as the use of full NPs enabled the participants to make sure that they were both talking about the same character (for a similar rationale on how low-level memory processes may positively impact communication, see Shintel & Keysar, 2009).

The method used in the current study also enabled us to examine the participants' referential choices as a function of their progress in the interaction and of the complexity of the story they were telling. Regarding the influence of the participants' progress in the interaction, a different pattern of

1
2
3 results was found for the two conditions (alternating or control). In the alternating condition, we found
4 no statistical evidence that the participants' choices to produce reduced forms to maintain a referent in
5 focus depended on where they were in the story. However, in the control condition, participants were
6 less likely to use a reduced form to maintain character 2 than to maintain character 1 (regardless of
7 whether character 1 was being maintained for the first time or the second time); which implies that the
8 director was more likely to produce a full NP to maintain character 2 in this condition. This finding
9 could be due to the director in the control condition paying particular attention to mutual
10 comprehension after the shift to character 2, in the middle of the story, particularly since the topical
11 status of this second character is not necessarily always clearly established (for a similar finding, see
12 Hendriks et al., 2014).

21
22 Why was the pattern of results different in both conditions? In the alternating condition, the
23 probability of producing a reduced form to maintain a referent remained relatively low throughout the
24 entire interaction (i.e., lower than in the control condition), possibly because the participants preferred
25 to also produce a full NP to check and signal that they were talking about the same character
26 previously introduced or reintroduced by the other participant. In contrast, in the control condition, the
27 participants did not need to engage into such checks after each speech turn; rather, it seems that the
28 directors optimized mutual comprehension by using less reduced forms, and hence more full NPs,
29 specifically for when he or she needed to refer to character 2 (maintaining of character 2 stage), thus
30 emphasizing the character change for the matcher. A similar result was also observed in Hendriks et
31 al.'s study (2014), who reported a more explicit referential behavior from their adult speakers, when
32 participants were maintaining reference to character 2, with a greater use of full NPs rather than
33 pronouns for this secondary character. Possible explanations included high sensitivity from speakers
34 to the possibility of the listener misunderstanding a reduced form such as a pronoun (see also
35 Contemori and Dussias, 2016), especially when referring to a less central character in the story. It thus
36 appears that depending on the features of the dialogue situation, mutual comprehension (Clark &
37 Brennan, 1991) may be reached in different ways. When referents are introduced and maintained by
38 the same person, this person attempts to facilitate their partner's comprehension by producing more
39 explicit information in situations where the risk of miscomprehension is higher (as after the shift to
40 character 2 in our stories). But when referents are introduced and maintained by two different people,
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 the risk of miscomprehension must be managed on a turn-by-turn basis, which results in an overall
4 decreased probability of producing reduced forms.
5

6
7 Finally, the method used also enabled us to examine the influence of referential complexity on
8 referential choices, as in the current study participants were instructed to tell stories involving one or
9 two characters. Analysis 1 replicated Fossard et al.'s (2018) finding that fewer pronouns and zero
10 anaphora were produced in stories which featured two characters (gender unambiguous and gender
11 ambiguous stories) than in stories which featured one character only. The difference between gender
12 unambiguous and gender ambiguous stories, which was non-significant in analysis 1 (as shown in
13 Figure 4, the likelihood of producing reduced forms was 0.69 in gender unambiguous stories and 0.51
14 in gender ambiguous stories, $p = .054$), reached significance in analysis 2 (which only included the
15 data from these two levels). This significant effect suggests that manipulating referential complexity
16 through character gender can also affect the speakers' choices to produce reduced referential forms.
17 This result is also in line with previous research, which indicated that manipulating the referent's
18 gender results in a decrease in the use of pronouns in gender ambiguous contexts (cf. Fukumura et
19 al., 2010; Contemori and Dussias, 2016). Interestingly, the percentage of reduced forms in the current
20 study remained below 70% (i.e., on average, the likelihood of producing reduced forms was below .70
21 in both gender unambiguous and gender ambiguous conditions), which is lower than what was found
22 by Fossard et al. (2018), who reported more than 85% of pronouns in both gender unambiguous and
23 gender ambiguous stories, but much higher than other studies (e.g., less than 20% of pronoun used in
24 gender unambiguous stories in Arnold and Griffin's study, 2007). These differences could be due to
25 the type of task used (story completion in Arnold and Griffin's (2007) study vs. storytelling in sequence
26 task in the current study) as well as changes in the procedure used here compared with Fossard et
27 al.'s (2018) initial study. In the current study, the participants alternated between trials in the control
28 condition and trials in the alternating condition; in contrast, there was no alternating condition in
29 Fossard et al.'s study. This could have affected reference production, for instance by making the
30 participants overall more cautious regarding the use of reduced forms. Fossard et al.'s study also
31 involved French-speaking participants, whereas the current study involved English-speaking
32 participants. Linguistic differences between the English and French referential systems may play a
33 role in the observed differences. What is more, in Fossard et al.'s study, only the referential
34 expressions in subject position (representing 98.15% of the data collected) were included in the
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 analysis; such restriction was not applied in the current study. More generally speaking, the fact that
4 the stories were divided into six images, as well as the fact that the participants did not see all of the
5 images to start with in the alternating condition could have altered the dynamics of reference
6 production, causing differences with Fossard et al.'s study in which participants could see all six
7 pictures in the correct order from the very beginning of each trial (see Canoz & Vion, 1994; Trabasso
8 & Nickels, 1992). Finally, we cannot discard the possibility that some of the differences observed were
9 due to Fossard et al.'s study involving only one naïve participant and a research assistant (who knew
10 the stories already), whereas the current study involved two naïve participants (see Kuhlen &
11 Brennan, 2013, for a discussion of how confederates who know "too much" about the task might
12 inadvertently convey this to the naïve participants, thus affecting the amount of information provided
13 by the latter during the interaction).

14
15 To summarize, our findings extend previous findings on referential choices in discourse
16 (Arnold, 2010; Arnold & Griffin, 2007; Colle et al., 2008; Hendriks et al., 2014) and dialogue (Brennan
17 & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Fossard et al., 2018; Knutsen et al., 2018) by showing that
18 dialogue partners are less likely to use reduced forms (and hence more likely to use full NPs) when a
19 referent is introduced by one person and maintained by someone else (our alternating condition) than
20 when all referential choices are made by the same person throughout the interaction (our control
21 condition).

22
23 The current work raises a number of new research questions. Firstly, the material was
24 constructed such that in the alternating condition, participants switched roles at specific moments in
25 the stories, one participant being responsible for the three maintaining images on which we focused
26 for the analyses and the other one being responsible for the introduction and reintroduction of the
27 characters (or additional maintaining images in simple stories for which the reintroduction stage did
28 not apply). Future work should examine referential choices in situations where dialogue partners
29 decide when to switch roles (speaker or listener) in order to determine whether similar patterns of
30 referential choices are found.

31
32 It is also noteworthy that because the participants' task was to enable their partner to find the
33 picture they were talking about, the increase in the production of reduced forms for picture 6 might
34 simply have been due to the task being complete by then (i.e., once all pictures had been placed, the
35 last picture left was necessarily picture 6). That being said, the participants were also instructed to *tell*

1
2
3 a story, which should have led them to produce referential forms which were readily understandable to
4
5 their partner. We thus believe that the results corresponding to picture 6 are *not* biased due to our
6
7 methodological choices. However, we will attempt to overcome this issue in our future work by giving
8
9 the matcher more than 6 cards (e.g., a 7th card featuring the same character(s) as in the rest of the
10
11 story may be used). This will prevent the director from assuming that once five pictures have been
12
13 placed, the matcher automatically knows which picture should be placed in final position.

14
15 Another question raised by this research pertains to the generalizability of our findings to real-
16
17 life dialogue situations, in which people are not usually explicitly instructed to take turns producing
18
19 referential expressions. The experimental setting used in the current study was designed to enable us
20
21 to test our hypotheses, but it prevented us from examining referential choices in situations where
22
23 people choose not only how to refer to things, but also when to refer to them. Future studies should
24
25 attempt to overcome this issue in order to offer a better understanding of the mechanisms which
26
27 underlie referential choices in more spontaneous dialogue settings.

28
29 Beyond the maintaining stage, future research should also examine other aspects of
30
31 referential choices, including choices regarding reference content (cf. Fossard et al., 2018). In
32
33 particular, we plan to investigate cases in which dialogue partners switch from using a full NP with a
34
35 modifier (e.g., “the little boy with a toy plane”) to a full NP without a modifier (e.g., “the boy”) or a full
36
37 NP with a shorter modifier (e.g., “the little boy”). This additional research will enable us to determine
38
39 how dialogue partners balance grounding constraints and individual cognitive costs associated with
40
41 producing longer references. For instance, switching from a longer to a shorter modifier might be a
42
43 good strategy to ground information explicitly while limiting the number of words produced.

43 **4.1. Conclusion**

44
45 All in all, we have shown that referential choices depend on *who says what* in the discourse setting.
46
47 Whereas research on narrative production by a single speaker suggests that references are usually
48
49 maintained using reduced forms (Arnold, 2010; Arnold & Griffin, 2007; Hendriks et al., 2014), the
50
51 current study shows that this effect is less apparent in settings where referential choices are managed
52
53 collectively. What is more, different dialogue features constrain referential choices depending on
54
55 whether speakers take turns producing reference or only one of the speakers produces all of the
56
57 references, highlighting the difference between referent maintaining in discourse settings and referent
58
59 maintaining in dialogic settings.
60

Acknowledgements

The authors thank Kinsey Goldman da Cunha, Cordetta Richmond and Niamh Meekings for their assistance in data collection and coding. A Swiss National Foundation (FNS) Grant (n°140269) to the second author supported the development of the materials used in this study. The third author is supported by a salary award from Fonds de recherche du Québec –Santé (FRQ-S).

Declarations of interest

None.

Author contributions

The first author conceptualized the study, supervised the data collection, transcription and analysis, and wrote the original draft. The second and third authors supported the development of the material and coding, participated to the conceptualization of the study, supervision of transcription, and writing of the original draft.

Peer Review Version

References

- Achim, A.M., Fossard, M., Couture, S., & Achim, A. (2015). Adjustment of speaker's referential expressions to an addressee's likely knowledge and link with theory of mind abilities. *Frontiers in Psychology*, 6:823. doi:103389/fpsyg.2015.00823
- Achim, A. M., Achim, A., & Fossard, M. (2017). Knowledge likely held by others affects speakers' choices of referential expressions at different stages of discourse. *Language, Cognition and Neuroscience*, 32, 21-36. <https://doi.org/10.1080/23273798.2016.1234059>
- Ariel, M. (1990). *Accessing noun-phrase antecedents*. Routledge.
- Ariel, M. (2001). Accessibility theory : An overview. In T. Sanders, J. Schilperoord, & W. Spooren (Éds.), *Text Representation : Linguistic and Psycholinguistic Aspects* (p. 29-87). John Benjamins Publishing Company.
- Arnold, J. E. (2010). How speakers refer : The role of accessibility. *Language and Linguistics Compass*, 4, 187-203. <https://doi.org/10.1111/j.1749-818X.2010.00193.x>
- Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression : Everyone counts. *Journal of Memory and Language*, 56, 521-536. <https://doi.org/10.1016/j.jml.2006.09.007>
- 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. <https://doi.org/10.1016/j.jml.2012.11.001>
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482-1493. <https://doi.org/10.1037/0278-7393.22.6.1482>
- Canoz, F., & Vion, M. (1994). Encoding and maintaining reference in oral discourse. *International Journal of Psychology*, 29, 537-564. <https://doi.org/10.1080/00207599408248173>
- Contemori, C., & Dussias, P.E. (2016). Referential choice in a second language: Evidence for a listener-oriented approach. *Language, Cognition and Neuroscience*, 31, 1257-1272. <https://doi.org/10.1080/23273798.2016.1220604>
- Clark, H. H., & Bangerter, A. (2004). Changing conceptions of reference. In I. Noveck & D. Sperber (Éds.), *Experimental pragmatics* (p. 25-49). Palgrave Macmillan.

- 1
2
3 Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, &
4 S. D. Teasley (Éds.), *Perspectives on socially shared cognition* (p. 127-149). American
5 Psychological Association.
6
7
8
9 Clark, H. H., & Marshall, C. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L.
10 Webber, & I. A. Sag (Éds.), *Elements of discourse understanding* (Vol. 2, p. 10-63).
11 Cambridge University Press.
12
13
14 Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13, 259-294.
15 https://doi.org/10.1207/s15516709cog1302_7
16
17
18 Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.
19 [https://doi.org/10.1016/0010-0277\(86\)90010-7](https://doi.org/10.1016/0010-0277(86)90010-7)
20
21
22 Colle, L., Baron-Cohen, S., Wheelwright, S., & Van der Lely, H. K. (2008). Narrative discourse in
23 adults with high-functioning autism or Asperger syndrome. *Journal of Autism and*
24 *Developmental Disorders*, 38, 28-40. <https://doi.org/10.1007/s10803-007-0357-5>
25
26
27
28 Fossard, M., Achim, A. M., Rousier-Vercruyssen, L., Gonzalez, S., Bureau, A., & Champagne-Lavau,
29 M. (2018). Referential choices in a collaborative storytelling task : Discourse stages and
30 referential complexity matter. *Frontiers in Psychology*, 9, 1-18.
31 <https://doi.org/10.3389/fpsyg.2018.00176>
32
33
34
35
36 Fox Tree, J. E., & Clark, N. B. (2013). Communicative Effectiveness of Written Versus Spoken
37 Feedback. *Discourse Processes*, 50(5), 339-359.
38 <https://doi.org/10.1080/0163853X.2013.797241>
39
40
41
42 Fukumura, K., van Gompel, R.P., and Pickering, M.J. (2010). The use of visual context during the
43 production of referring expressions. *Quarterly Journal of Experimental Psychology*, 63, 1700-
44 1715. <https://doi:10.1080/17470210903490969>
45
46
47
48 Gatt, A., Krahmer, E., van Deemter, K., & van Gompel, R. P. G. (2014). Models and empirical data for
49 the production of referring expressions. *Language, Cognition and Neuroscience*. 29, 899-911.
50 <https://doi.org/10.1080/23273798.2014.933242>
51
52
53
54 Gergle, D., Kraut, R. E., & Fussell, S. R. (2013). Using visual information for grounding and awareness
55 in collaborative tasks. *Human-Computer Interaction*, 28, 1-39.
56 <https://doi.org/10.1080/07370024.2012.678246>
57
58
59 Givon, T. (1983). *Topic continuity in discourse*. John Benjamins.
60

- 1
2
3 Goudbeek, M., & Krahmer, E. (2012). Alignment in interactive reference production : Content planning,
4
5 modifier ordering and referential overspecification. *Topics in Cognitive Science*, 4, 269-289.
6
7 <https://doi.org/10.1111/j.1756-8765.2012.01186.x>
8
9 Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering : A framework for modelling the local
10
11 discourse. *Computational Linguistics*, 21, 203-225. <https://doi.org/10.21236/ADA324949>
12
13 Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring
14
15 expressions in discourse. *Language*, 69, 274-307. <https://doi.org/10.2307/416535>
16
17 Hendriks, P., Koster, C., & Hoeks, J. C. (2014). Referential choice across the lifespan : Why children
18
19 and elderly adults produce ambiguous pronouns. *Language, Cognition and Neuroscience*, 29,
20
21 391-407. <https://doi.org/10.1080/01690965.2013.766356>
22
23 Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design : Knowing when
24
25 and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, 47,
26
27 589-606. [https://doi.org/10.1016/S0749-596X\(02\)00019-0](https://doi.org/10.1016/S0749-596X(02)00019-0)
28
29 Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*,
30
31 59, 91-117. [https://doi.org/10.1016/0010-0277\(96\)81418-1](https://doi.org/10.1016/0010-0277(96)81418-1)
32
33 Ibarra, A., & Tanenhaus, M. K. (2016). The flexibility of conceptual pacts : Referring expressions
34
35 dynamically shift to accommodate new conceptualizations. *Frontiers in Psychology*.
36
37 <https://doi.org/10.3389/fpsyg.2016.00561>
38
39 Kiernan, K., Tao, J., & Gibbs, P. (2012). *Tips and strategies for mixed modelling with SAS/STAT*
40
41 *procedures*. 2012 SAS Global Forum, Orlando, FL.
42
43 Knutsen, D., & Le Bigot, L. (2012). Managing dialogue : How information availability affects
44
45 collaborative reference production. *Journal of Memory and Language*, 67, 326-341.
46
47 <https://doi.org/10.1016/j.jml.2012.06.001>
48
49 Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative
50
51 dialogue. *Psychonomic Bulletin & Review*, 21, 1590-1599. [https://doi.org/10.3758/s13423-014-](https://doi.org/10.3758/s13423-014-0620-7)
52
53 0620-7
54
55 Knutsen, D., & Le Bigot, L. (2015). The influence of reference acceptance and reuse on
56
57 conversational memory traces. *Journal of Experimental Psychology: Learning, Memory, and*
58
59 *Cognition*, 41, 574-585. <https://doi.org/10.1037/xlm0000036>
60

- 1
2
3 Knutsen, D., Ros, C., & Le Bigot, L. (2018). Spoilt for choice : Initially considering several referential
4 expressions affects subsequent referential decisions. *Language, Cognition and Neuroscience*,
5 33, 618-632. <https://doi.org/10.1080/23273798.2017.1400080>
6
7
8
9 Koolen, R., Gatt, A., Goudbeek, M., & Kramer, E. (2011). Factors causing overspecification in definite
10 description. *Journal of Pragmatics*, 43 ,3231-3250.
11
12 <https://doi.org/10.1016/j.pragma.2011.06.008>
13
14
15 Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue : When confederates might be
16 hazardous to your data. *Psychonomic Bulletin & Review*, 20, 54-72.
17
18 <https://doi.org/10.3758/s13423-012-0341-8>
19
20
21 Lysander, K., & Horton, W. S. (2012). Conversational Grounding in Younger and Older Adults : The
22 Effect of Partner Visibility and Referent Abstractness in Task-Oriented Dialogue. *Discourse*
23 *Processes*, 49(1), 29-60. <https://doi.org/10.1080/0163853X.2011.625547>
24
25
26 McInnes, F., & Attwater, D. (2004). Turn-taking and grounding in spoken telephone number transfers.
27 *Speech Communication*, 43, 205-223. <https://doi.org/10.1016/j.specom.2004.04.001>
28
29
30 Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and*
31 *Brain Sciences*, 27, 169-190. <https://doi.org/10.1017/S0140525X04000056>
32
33
34 Roque, A., & Traum, D. (2008). Degrees of grounding based on evidence of understanding.
35 *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, 54-63.
36
37
38 Rosnagel, C. (2000). Cognitive load and perspective-taking : Applying the automatic-controlled
39 distinction to verbal communication. *European Journal of Social Psychology*, 30, 429-445.
40
41 [https://doi.org/10.1002/\(SICI\)1099-0992\(200005/06\)30:3<429::AID-EJSP3>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1099-0992(200005/06)30:3<429::AID-EJSP3>3.0.CO;2-V)
42
43
44 Shintel, H., & Keysar, B. (2009). Less is more : A minimalist account of joint action in communication.
45 *Topics in Cognitive Science*, 1, 260-273. <https://doi.org/10.1111/j.1756-8765.2009.01018.x>
46
47
48 Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a
49 picture story. *Discourse Processes*, 15, 249-275. <https://doi.org/10.1080/01638539209544812>
50
51
52 Viethen, J., Dale, R., & Guhe, M. (2014). Referring in dialogue : Alignment or construction? *Language,*
53 *Cognition and Neuroscience*, 29, 950-974. <https://doi.org/10.1080/01690965.2013.827224>
54
55
56 Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation.
57 *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 919-937.
58
59 <https://doi.org/10.1037/a0036161>
60

Figure Captions

Figure 1. Example of stories used in Fossard et al.'s (2018) experiment and in the current experiment. The colored numbers are used for description purposes only (see below) and were not visible to the participants in the current experiment.

Figure 2. Grid and loose cards provided to the matcher in the control condition.

Figure 3. Grids and loose cards provided to the two participants in the alternating condition.

Figure 4. Probability of producing a reduced form during the maintaining stage as a function of condition and referential complexity. The error bars represent the standard error. *Note: The number of data points was 249 in the control – single character condition, 150 in the control – gender unambiguous condition, 152 in the control – gender ambiguous condition, 250 in the alternating – single character condition, 150 in the alternating – gender unambiguous condition and 147 in the alternating – gender ambiguous condition.*

Figure 5. Probability of producing a reduced form during the maintaining stages as a function of maintaining context and condition for gender unambiguous stories (upper panel) and gender ambiguous stories (lower panel). *Note: The number of data points was 50 in all conditions in the gender unambiguous stories; as for gender ambiguous stories, the number of data points was 51 in the control – first maintaining of character 1 condition, 51 in the control – first maintaining of character 2, 50 in the control – second maintaining of character 1 condition, 49 in the alternating – first maintaining of character 1 condition, 49 in the alternating – first maintaining of character 2 condition and 49 in the alternating – second maintaining of character 1 condition.*

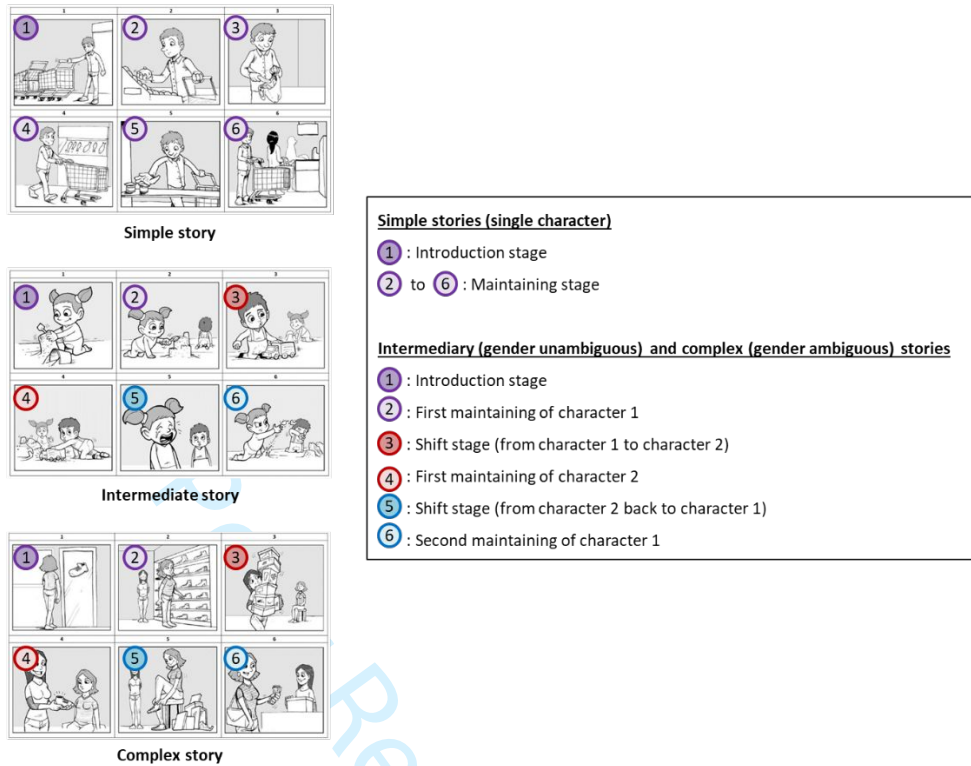


Figure 1. Example of stories used in Fossard et al.'s (2018) experiment and in the current experiment. The colored numbers are used for description purposes only (see below) and were not visible to the participants in the current experiment.

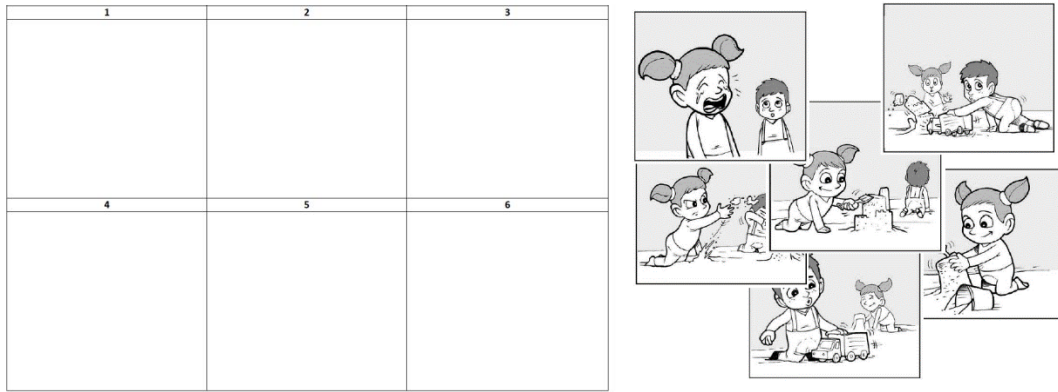


Figure 2. Grid and loose cards provided to the matcher in the control condition.

Peer Review Version

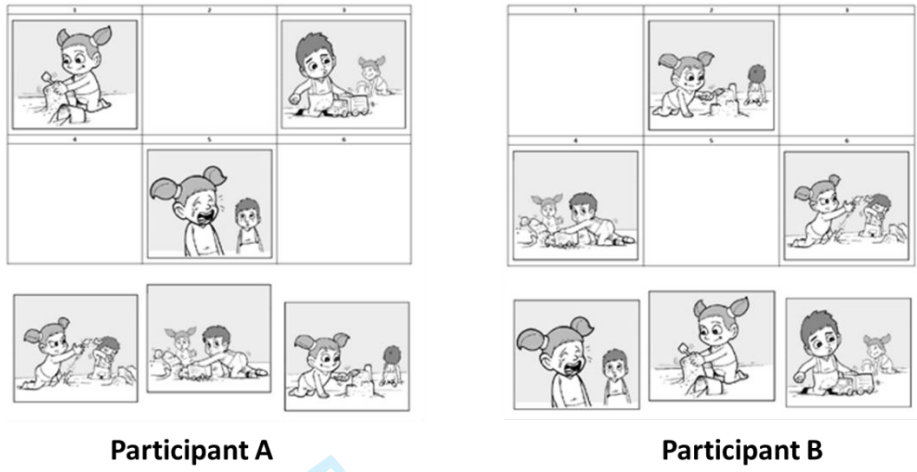


Figure 3. Grids and loose cards provided to the two participants in the alternating condition.

Peer Review Version

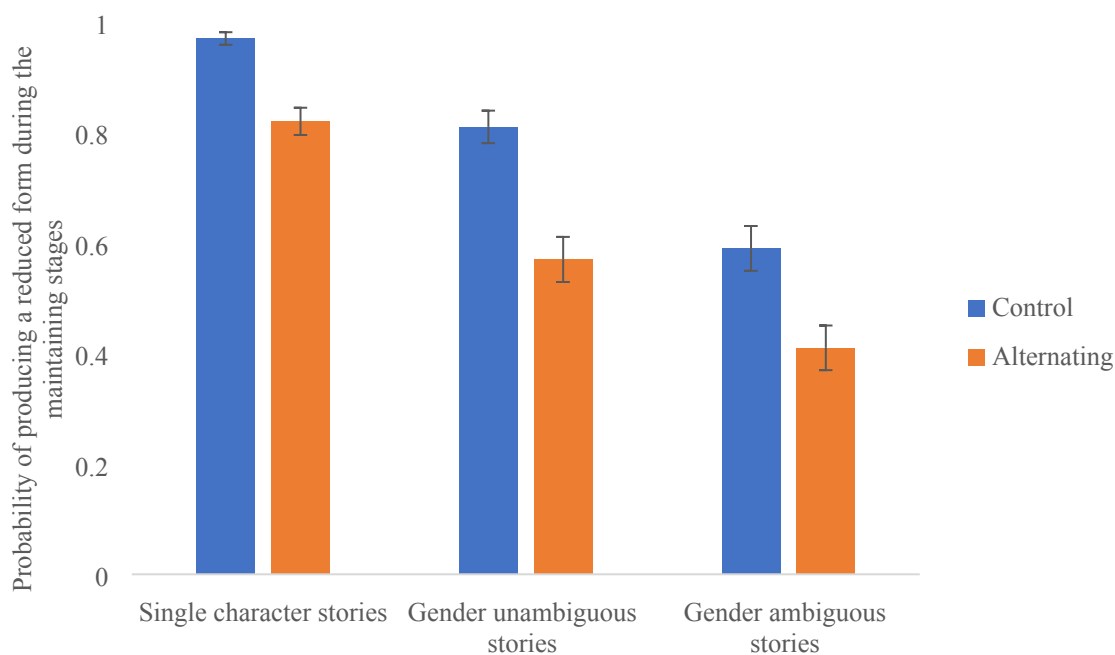


Figure 4. Probability of producing a reduced form during the maintaining stage as a function of condition and referential complexity. The error bars represent the standard error. *Note:* The number of data points was 249 in the control – single character condition, 150 in the control – gender unambiguous condition, 152 in the control – gender ambiguous condition, 250 in the alternating – single character condition, 150 in the alternating – gender unambiguous condition and 147 in the alternating – gender ambiguous condition.

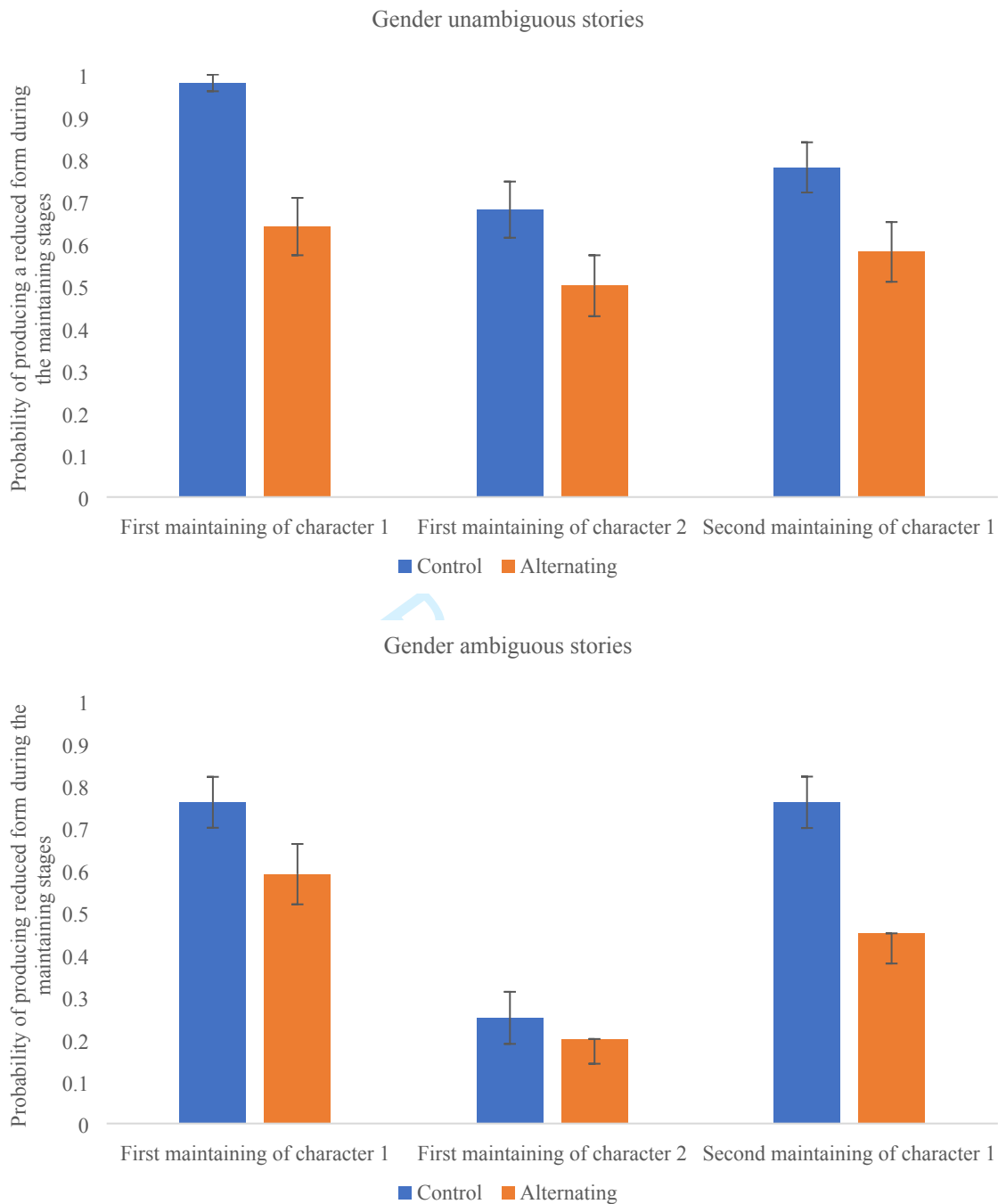


Figure 5. Probability of producing a reduced form during the maintaining stages as a function of maintaining context and condition for gender unambiguous stories (upper panel) and gender ambiguous stories (lower panel). *Note:* The number of data points was 50 in all conditions in the gender unambiguous stories; as for gender ambiguous stories, the number of data points was 51 in the control – first maintaining of character 1 condition, 51 in the control – first maintaining of character 2, 50 in the control – second maintaining of character 1 condition, 49 in the alternating – first

1
2
3 *maintaining of character 1 condition, 49 in the alternating – first maintaining of character 2 condition*
4
5 *and 49 in the alternating – second maintaining of character 1 condition.*
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Peer Review Version

Table 1

Dialogue Sample – Control Condition

Dyad	Condition	Complexity	Picture	Picture type	Director	Director speech
10	Control	Single character	1	Introduction	A	a man grabs a shopping trolley
10	Control	Single character	2	Maintaining of character 1	A	he then proceeds to go shopping and starts picking out potatoes
10	Control	Single character	3	Maintaining of character 1	A	he then places cabbages in his bag no sorry cucumbers in his bag sorry
10	Control	Single character	4	Maintaining of character 1	A	he then is walking with the shopping trolley past fish
10	Control	Single character	5	Maintaining of character 1	A	he then grabs some yoghurt
10	Control	Single character	6	Maintaining of character 1	A	and then he's all done and is about to buy it
12	Control	Gender unambiguous	1	Introduction	B	there's a little girl on the beach building a sandcastle
12	Control	Gender unambiguous	2	First maintaining of character 1	B	the sandcastle's nearly built and there's a boy in the distance
12	Control	Gender unambiguous	3	Shift to character 2	B	the boy's playing with a truck or something
12	Control	Gender unambiguous	4	First maintaining of character 2	B	and then he knocks into her sandcastle and knocks it down
12	Control	Gender unambiguous	5	Shift back to character 1	B	and she starts to cry
12	Control	Gender unambiguous	6	Second maintaining of character 1	B	and puts sand in his face
14	Control	Gender ambiguous	1	Introduction	A	a lady is walking past a shoe shop with a shoe in the window that catches her attention
14	Control	Gender ambiguous	2	First maintaining of character 1	A	she enters the shop and is looking at a wall with different shoes on it
14	Control	Gender ambiguous	3	Shift to character 2	A	she can't decide what ones she wants so the shop assistant brings out loads of boxes for her
14	Control	Gender ambiguous	4	First maintaining of character 2	A	the shop assistant presents her with one shoe looking really happy
14	Control	Gender ambiguous	5	Shift back to character 1	A	the lady's trying on the shoe
14	Control	Gender ambiguous	6	Second maintaining of character 1	A	and then she's happy with it so she's paying for it

Note. The examples provided match the stories shown in Figure 1.

Table 2

Dialogue Sample – Alternating Condition

Dyad	Condition	Complexity	Picture	Picture type	Director	Director speech
20	Alternating	Single character	1	Introduction	A	the person is putting a trolley back at the supermarket
20	Alternating	Single character	2	Maintaining of character 1	B	so he's in the supermarket and he's holding a potato or something
20	Alternating	Single character	3	Maintaining of character 1	A	he's putting like a cucumber into his bag
20	Alternating	Single character	4	Maintaining of character 1	B	so he's passing by the fish stall or something
20	Alternating	Single character	5	Maintaining of character 1	A	he's putting like yogurt onto a trolley
20	Alternating	Single character	6	Maintaining of character 1	B	and he's in the queue behind a customer
17	Alternating	Gender unambiguous	1	Introduction	A	okay so you have a little girl that's erm making a sandcastle by herself and her bucket is tipped over
17	Alternating	Gender unambiguous	2	First maintaining of character 1	B	you then see her like making her sandcastle bigger with a spade and there's like a little boy in the background
17	Alternating	Gender unambiguous	3	Shift to character 2	A	the next one is a little boy he's got his tractor and he's just kinda playing with his little not a tractor like a van and then she's still making her castle her sandcastle in the background
17	Alternating	Gender unambiguous	4	First maintaining of character 2	B	and then you see him like running over her sandcastle with his van
17	Alternating	Gender unambiguous	5	Shift back to character 1	A	and then she starts crying cause now her sandcastle's just destroyed
17	Alternating	Gender unambiguous	6	Second maintaining of character 1	B	and then you see her like throwing sand at him
10	Alternating	Gender ambiguous	1	Introduction	B	a woman is notices there is a shoe shop and she's about to walk in
10	Alternating	Gender ambiguous	2	First maintaining of character 1	A	she stands inside looking at the shoes on offer as the store clerk stands beside her smiling
10	Alternating	Gender ambiguous	3	Shift to character 2	B	the store clerk brings in six or seven different shoes for her to try on
10	Alternating	Gender ambiguous	4	First maintaining of character 2	A	she takes one of the shoes out of the box and presents it to her in her hands

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

10	Alternating	Gender ambiguous	5	Shift back to character 1	B	the woman puts on the shoe and she seems to like it
10	Alternating	Gender ambiguous	6	Second maintaining of character 1	A	before taking it off and paying for the shoe at the front of the store

Note. The examples provided match the stories shown in Figure 1.

Peer Review Version

Table 3

Coding Table Extract – Control Condition

Dyad	Condition	Complexity	Picture	Reference	Reduced form
10	Control	Single character	2	<i>He</i>	Yes
10	Control	Single character	3	<i>He</i>	Yes
10	Control	Single character	4	<i>He</i>	Yes
10	Control	Single character	5	<i>He</i>	Yes
10	Control	Single character	6	<i>He</i>	Yes
12	Control	Gender unambiguous	2	<i>NA</i>	NA
12	Control	Gender unambiguous	4	<i>He</i>	Yes
12	Control	Gender unambiguous	6	\emptyset	Yes
14	Control	Gender ambiguous	2	<i>She</i>	Yes
14	Control	Gender ambiguous	4	<i>the shop assistant</i>	No
14	Control	Gender ambiguous	6	<i>She</i>	Yes

Note. NA = non-applicable. \emptyset = zero anaphora

Table 4

Coding Table Extract – Alternating Condition

Dyad	Condition	Complexity	Picture	Reference	Reduced form
20	Alternating	Single character	2	<i>He</i>	Yes
20	Alternating	Single character	3	<i>He</i>	Yes
20	Alternating	Single character	4	<i>He</i>	Yes
20	Alternating	Single character	5	<i>He</i>	Yes
20	Alternating	Single character	6	<i>He</i>	Yes
17	Alternating	Gender unambiguous	2	<i>Her</i>	Yes
17	Alternating	Gender unambiguous	4	<i>Him</i>	Yes
17	Alternating	Gender unambiguous	6	<i>Her</i>	Yes
10	Alternating	Gender ambiguous	2	<i>She</i>	Yes
10	Alternating	Gender ambiguous	4	<i>She</i>	Yes
10	Alternating	Gender ambiguous	6	∅	Yes

Note. ∅= zero anaphora

Table 5

Coefficients Associated with the Random Effects included in Analysis 1

Model parameter	Estimate	Standard error
By-dyad random intercepts	0.46	0.35
By-dyad random slopes corresponding to condition	0.21	0.26
By-participant random intercepts	0.33	0.30
By-participant random slopes corresponding to complexity	0.19	0.17
By-participant random slopes corresponding to condition	0.30	0.29
By-item random intercepts	0.11	0.12
By-item random slopes corresponding to condition	0.03	< 0.01

Peer Review Version

Table 6

Coefficients Associated with the Fixed Effects included in Analysis 1

Effect	Estimate	Standard error	95% CI (lower bound)	95% CI (upper bound)
Intercept	3.75	0.47	2.82	4.69
Complexity: Gender ambiguous	-3.37	0.50	-4.39	-2.35
Complexity: Gender unambiguous	-2.08	0.52	-3.13	-1.04
Complexity: Single character (baseline)	0			
Condition: Alternating	-1.89	0.46	-2.79	-0.99
Condition: Control (baseline)	0			
Complexity: Gender ambiguous x Condition: Alternating	1.05	0.50	0.06	2.03
Complexity: Gender ambiguous x Condition : Control (baseline)	0			
Complexity: Gender unambiguous x Condition : Alternating	0.52	0.52	-0.50	1.53
Complexity: Gender unambiguous x Condition: Control (baseline)	0			
Complexity: Single character x Condition: Alternating (baseline)	0			
Complexity: Single character x Condition: Control (baseline)	0			

Table 7

Coefficients Associated with the Random Effects included in Analysis 2

Model parameter	Estimate	Standard error
By-dyad random intercepts	0.05	0.37
By-participant random intercepts	0.95	0.55
By-participant random slopes corresponding to complexity	0.02	0.22
By-participant random slopes corresponding to condition	0.49	0.33
By-item random intercepts	0.05	0.23
By-item random slopes corresponding to maintaining context	0.52	0.31

Peer Review Version

Table 8

Coefficients Associated with the Fixed Effects included in Analysis 2

Effect	Estimate	Standard error	95% CI (lower bound)	95% CI (upper bound)
Intercept	1.55	0.57	0.39	2.71
Complexity level: Gender ambiguous	-0.16	0.76	-1.71	1.40
Complexity level: Gender unambiguous (baseline)	0			
Condition: Alternating	-1.15	0.54	-2.21	-0.10
Condition: Control (baseline)	0			
Complexity: Gender ambiguous x Condition: Alternating	-0.51	0.73	-1.93	0.92
Complexity: Gender ambiguous x Condition: Control (baseline)	0			
Complexity: Gender unambiguous x Condition: Alternating (baseline)	0			
Complexity: Gender unambiguous x Condition: Control (baseline)	0			
Context: 1 st maintaining of C1	2.82	1.22	0.42	5.22
Context: 1 st maintaining of C2	-0.71	0.72	-2.21	0.79
Context: 2 nd maintaining of C1 (baseline)	0			
Context: 1 st maintaining of C1 x Complexity: Gender ambiguous	-2.82	1.42	-5.64	< -0.01
Context: 1 st maintaining of C1 x Complexity: Gender unambiguous (baseline)	0			
Context: 1 st maintaining of C2 x Complexity: Gender ambiguous	-2.00	1.02	-4.12	0.12
Context: 1 st maintaining of C2 x Complexity: Gender unambiguous (baseline)	0			
Context: 2 nd maintaining of C1 x Complexity: Gender ambiguous (baseline)	0			
Context: 2 nd maintaining of C1 x Complexity: Gender unambiguous (baseline)	0			
Context: 1 st maintaining of C1 x Condition: Alternating	-2.54	1.20	-4.90	-0.18
Context: 1 st maintaining of C1 x Condition: Control (baseline)	0			
Context: 1 st maintaining of C2 x Condition: Alternating	0.29	0.69	-1.07	1.65
Context: 1 st maintaining of C2 x Condition: Control (baseline)	0			
Context: 2 nd maintaining of C1 x Condition: Alternating	0			
Context: 2 nd maintaining of C1 x Condition: Control	0			
Context: 1 st maintaining of C1 x Complexity: Gender ambiguous x Condition: Alternating	3.30	1.40	0.55	6.04
Context: 1 st maintaining of C1 x Complexity: Gender ambiguous x Condition: Control (baseline)	0			
Context: 1 st maintaining of C1 x Complexity: Gender unambiguous x Condition: Alternating (baseline)	0			
Context: 1 st maintaining of C1 x Complexity: Gender unambiguous x Condition: Control (baseline)	0			

1					
2					
3	Context: 1 st maintaining of C2 x Complexity:	1.00	1.00	-0.97	2.97
4	Gender ambiguous x Condition: Alternating				
5	Context: 1 st maintaining of C2 x Complexity:	0			
6	Gender ambiguous x Condition: Control				
7	(baseline)				
8	Context: 1 st maintaining of C2 x Complexity:	0			
9	Gender unambiguous x Condition:				
10	Alternating (baseline)				
11	Context: 1 st maintaining of C2 x Complexity:	0			
12	Gender unambiguous x Condition: Control				
13	(baseline)				
14	Context: 2 nd maintaining of C1 x Complexity:	0			
15	Gender ambiguous x Condition: Alternating				
16	(baseline)				
17	Context: 2 nd maintaining of C1 x Complexity:	0			
18	Gender ambiguous x Condition: Control				
19	(baseline)				
20	Context: 2 nd maintaining of C1 x Complexity:	0			
21	Gender unambiguous x Condition:				
22	Alternating (baseline)				
23	Context: 2 nd maintaining of C1 x Complexity:	0			
24	Gender unambiguous x Condition: Control				
25	(baseline)				
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					
49					
50					
51					
52					
53					
54					
55					
56					
57					
58					
59					
60					