Running head: Reusing preferred referential expressions

Spoilt for choice: Initially considering several referential expressions affects subsequent referential decisions

Dominique KNUTSEN

Department of Psychology, University of Essex, UK

Christine ROS and Ludovic LE BIGOT

Université de Poitiers, France; CNRS (CeRCA UMR 7295), France

Corresponding author:

Dr Dominique Knutsen

University of Essex

Wivenhoe Park

Colchester CO4 3SQ

United Kingdom

Email address: dknutsen@essex.ac.uk

Abstract

In dialogue, speakers jointly decide how to refer to the referents under discussion. In some cases, several different referential expressions are considered before the partners can decide which one they prefer; this work examined how doing so affects subsequent referential expression reuse. Pairs of participants came up with suitable referential expressions for Tangram figures they were shown. They then referred to the same figures again during a matching task which was performed either with the same partner or a different partner. The main finding was that the preferred referential expression was less likely to be reused when several referential expressions were initially considered. This effect could not be attributed to a generation effect or to some referential expressions being a better match for the Tangram figures than others. These findings offer a better understanding of how the initial contribution of a reference shapes subsequent referential decisions through ordinary memory functioning.

Keywords

Dialogue; common ground; memory; contribution; referential decisions

Word count (excluding acknowledgements, reference list and appendices)

8,873

Spoilt for choice: Initially considering several referential expressions affects subsequent referential decisions

Introduction

Dialogue is a joint activity in which at least two partners collaborate in order to achieve a common goal (e.g., scheduling a meeting or planning a holiday trip; Bangerter & Clark, 2003; Bangerter, Clark, & Katz, 2004; Clark, 1992, 1996; Mills, 2014; Tylén, Weed, Wallentin, Roepstorff, & Frith, 2010). One fundamental feature of this activity is that it requires speakers to make decisions about how to refer to the objects and entities under discussion, as there are usually several different ways of referring to the same thing. For instance, the same animal might be referred to as "a cat", "a kitty" or "an adorable ball of fur" (Brennan & Clark, 1996; Clark & Bangerter, 2004; Clark & Marshall, 1981; Gorman, Gegg-Harrison, Marsh, & Tanenhaus, 2013; Horton & Gerrig, 2005a, 2016; Isaacs & Clark, 1987; Yoon & Brown-Schmidt, 2014). Initial referential decisions about how to refer to things are made collaboratively, with both partners playing an active role in the process (Clark & Schaefer, 1989). In some cases, the partners only consider one referential expression, which they both agree upon; in other cases, several different referential expressions are considered before the partners decide which referential expression they prefer. Both situations are not usually distinguished in experimental research (except for Tolins, Zeamer, & Fox Tree, 2017). The purpose of the current study was to show that initially considering several referential expressions leads to the construction of a less rich memory representation of the preferred referential expression, making the latter less likely to be reused subsequently.

The role of linguistic precedents in dialogic referential decisions

An important determinant of referential decisions is whether or not the referent under discussion has already been mentioned previously. Indeed, referential choices depend at least in part on the accessibility in memory of referential expressions: the more readily accessible a referential expression, the more likely it is to be produced by the dialogue partners (see Horton & Gerrig, 2005a, 2016). Accordingly, the presence of a suitable referential expression in the dialogue history increases its accessibility in the partners' dialogue memory, thus also increasing the likelihood that it will be reused later (for an example, see Rossnagel, 2000).

The fact that memory plays a central role in referential decisions has led dialogue researchers to investigate further which features of the interaction setting might contribute to making some referential expressions more readily accessible in memory than others. In particular, Horton and colleagues (Horton, 2007, 2008; Horton & Gerrig, 2005a, 2005b, 2016; Horton & Slaten, 2012; see also Barr, Jackson, & Phillips, 2014; Knutsen, Ros, & Le Bigot, 2016) have shown that when a referential expression is mentioned during an interaction, each person encodes this referential expression in association with the referent and his or her current dialogue partners. When encountering these partners again, the referential expression automatically becomes readily accessible in memory through resonance (see Ratcliff, 1978), thus making it more likely to be reused to refer to the same referent again during the current interaction.

Importantly, associations in memory between partners, referential expressions and referents are more or less strong. For instance, if a given referential expression is repeated several times during an interaction to refer to a given referent in the presence of a given dialogue partner, the resulting memory associations are stronger than if this referential expression was only mentioned once (Horton, 2008). In a similar way, Gorman et al. (2013) have reported that having the opportunity to learn referential expressions along with a dialogue partner (instead of learning these referential expressions individually and then being

told that one's dialogue partner learnt the same referential expressions) leads to the construction of richer memory representations, reinforcing partner-referential expression associations in memory, and thus making these referential expressions more likely to be reused subsequently.

These findings illustrate how low-level memory processes can lead speakers to favour the reuse of referential expressions which can be understood easily by their current partners. Indeed, if Speaker A reuses a referential expression which has already been used during a past interaction with Listener B, B should be capable of understanding this referential expression correctly. In other words, low-level processes lead speakers to (incidentally) favour the reuse of referential expressions which belong to the partners' *common ground*, which includes the knowledge that two dialogue partners share and are aware of sharing (e.g., Brown-Schmidt, 2012; Clark, 1996; Clark & Marshall, 1981).

The research reviewed above suggests that the more readily accessible in memory a referential expression, the more likely it is to be subsequently reused. It is noteworthy that although the memory-based approach to dialogue described above emphasises that memory associations between referents, referential expressions and dialogue partners play a central role in referential decisions, most – if not all – recent research about this topic has focused on the factors which affect the strength of memory associations between referential expressions and specific dialogue partners (e.g., Gorman et al., 2013; Horton & Gerrig, 2005b). The current study sought to take this work further by examining factors which might affect the strength of memory associations between specific referential expressions, specific dialogue partners *and specific referents*. The next section focuses on collaboration at the time of initial referential expression production and explores its potential influence on subsequent reuse.

Collaborative contributions to referential expression production in dialogue

When a referent is mentioned for the first time during an interaction, the dialogue partners must decide *together* how to refer to it. According to the contribution model (Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986), this is achieved in two phases. One of the partners starts by *presenting* a referential expression which he or she thinks his or her partner is capable of understanding; the other partner then *accepts* this referential expression by signalling that he or she believes that the referential expression presented was understood well enough for current purposes. Acceptance is more or less immediate, and more or less explicit, depending on the level of knowledge of the partner performing the acceptance. The contribution process enables dialogue partners to add referential expressions to their common ground. Once a referential expression has been added to the partners' common ground, or has been *grounded*, the participants can pursue the interaction. The contribution process is illustrated in the three examples below, in which A and B are looking at shoes in a shop (A and B are hereafter referred to as female and male, respectively; the example used is adapted from Brennan & Clark, 1996).

Example 1	**Example 2**	**Example 3**
A: I like the penny loafers. B: I think they're quite nice too.	A: I like the penny loafers. B: What's a penny loafer? A: These shoes here. B: I think they're quite nice too.	A: I like the penny loafers. B: Wouldn't you call those slip-on shoes, not penny loafers? A: The slip-on shoes, yes, that's right. B: I think they're quite nice too.

In Example 1, A presents the referential expression "the penny loafers" to refer to a pair of shoes she likes. A might have chosen this referential expression to distinguish these shoes from other kinds of shoes also visible in the shop at the time of the interaction (e.g. trainers and high-heel shoes). This referential expression is then accepted by B in the following speech turn, in which B implies that he has understood the referential expression presented by A. In this case, B's acceptance was implicit (i.e., B simply initiated the next

relevant speech turn after A's presentation). B could have accepted the same referential expression more explicitly, for instance by repeating it, by saying "mhm" or "ok" or by nodding his head (see Clark & Brennan, 1991; Clark & Krych, 2004; Fox Tree, 2010; McInnes & Attwater, 2004).

In Example 2, the referential expression presented by A cannot be accepted immediately by B, because B does not understand it. This leads B to initiate a repair process during which he asks A for additional information about her initial intention. B only accepts the referential expression presented initially once he has received enough information to understand it correctly.

Finally, in some cases, a new referential expression might be substituted to the referential expression initially presented, as illustrated in Example 3, in which B comes up with a referential expression he believes to be more appropriate than the one presented by A. Specifically, in this case, B signals that the referential expression presented initially is not suitable by presenting a different referential expression (i.e., "the slip-on shoes"). The content of A's following utterance suggests that she finds this new referential expression appropriate and thus that it can be accepted. In sum, in this kind of situation, two referential expressions are considered during the contribution process; the two partners actively take part in the decision as to which is most appropriate (this kind of referential expression is hereafter referred to as the *preferred* referential expression), highlighting the highly collaborative nature of the contribution process (Clark & Wilkes-Gibbs, 1986).

How might the number of referential expressions considered during the initial contribution process (i.e., one or more than one) affect subsequent reuse? The general hypothesis tested in the current study is that considering a single referential expression upon which both partners agree leads to the creation of a strong association between the referent and the chosen referential expression in the partners' memory. This, in turn, would increase

the likelihood of reusing this referential expression in subsequent interactions. In contrast, considering more than one referential expression would lead to the creation of a weaker association between the referent and the preferred referential expression in the partners' memory. This, in turn, would *decrease* the likelihood of reusing this referential expression in subsequent interactions.

In addition, recall that referents and that the corresponding (preferred or dispreferred) referential expressions are not only associated with each other in memory; they are also associated with information about the identity of the dialogue partners who were present when the referential expressions were initially considered (see Horton, 2007; Horton & Gerrig, 2005b). The presence of the initial dialogue partner might be sufficient to increase the level of accessibility of an associated referential expression, as the presence of two memory cues (i.e., the dialogue partner and the referent) might compensate for a weak memory association between a referent and the corresponding referential expression. Accordingly, the second general hypothesis tested in this study was that the effect of the number of referential expressions initially considered is stronger when referring repeatedly to a referent in the presence of a new dialogue partner than in the presence of the initial dialogue partner.

Overview and rationale of the experiment

An experiment was conducted in order to test these two hypotheses. The experiment was divided into two phases. During the first phase (Dialogue Phase), two participants had the opportunity to discuss abstract Tangram figures similar to that shown in Figure 1. Their task was to agree upon a label to refer to each picture. The participants then embarked on the second phase of the experiment (Matching Phase). In each trial of this phase, one of the participants (the Director) described a target Tangram figure to the other participant (the Matcher) so that the latter could identify the target among other Tangram figures. In critical

trials, the target had already been discussed during the Dialogue Phase. Pairs of participants took part in one of two conditions. In the "Same Partner" Condition, each participant performed the two phases with the same partner. In the "Different Partner" Condition, each participant switched partners between the Dialogue Phase and the Matching Phase.

Figure 1. Example of one of the Tangram figures used in this study.

Experiments using similar setups have found that the common ground built during the Dialogue Phase (i.e., the referential expressions presented and accepted during this phase) increases the partners' efficiency during the Matching Phase in the "Same Partner" Condition, because both partners favour the reuse of the referential expressions chosen previously to describe the target figures. In particular, this is reflected by a reduction of the number of words (content words and hedges such as *kind of*, *whatever*, *something* or *like*; see Brennan & Ohaeri, 1999; Liu & Fox Tree, 2012) produced by the partners in trials where referential expressions are available in the common ground (Clark & Wilkes-Gibbs, 1986; Hupet & Chantraine, 1992; Hupet, Chantraine, & Nef, 1993; Hupet, Seron, & Chantraine, 1991; Isaacs & Clark, 1987; Wilkes-Gibbs & Clark, 1992). Preliminary analyses of the data sought to confirm that the corpus gathered in this experiment exhibited these features.

Moreover, central to the rationale of the study was the fact that in some of the trials of the Dialogue Phase, one participant presented a referential expression and the other

participant came up with another, different referential expression; in some cases, the partners even came up with more than two referential expressions. Because the participants' task was to come up with a single label for each Tangram figure, this meant that the participants had to decide which referential expression they *preferred* and which referential expression(s) they *dispreferred*. In contrast, in trials where only one referential expression was considered by the participants, this single referential expression was necessarily the participants' preferred referential expression. The first operational hypothesis was that in the Matching Phase, Directors are more likely to reuse the preferred referential expression when the figure to describe had previously been associated with one referential expression only during the Dialogue Phase than when it has been associated with more than one referential expression. The second operational hypothesis was that this difference is stronger in the "Different Partner" Condition than in the "Same Partner" Condition. Indeed, as mentioned previously, the presence of the initial dialogue partner (in the "Same Partner" Condition) might be sufficient to increase the level of accessibility of a referential expression, even in cases where several referential expressions were initially considered.

Importantly, another factor which may also interfere with the reuse of the partners' preferred referential expression is who initially generated it, with each dialogue partner being more likely to reuse self-generated referential expressions than partner-generated ones (Knutsen & Le Bigot, 2014; Knutsen et al., 2016; Slamecka & Graf, 1978). This could be due to the fact that self-produced words are more readily accessible in memory than partner-produced words, making the former more likely to be produced (MacLeod, 2011). This could also be due to the participants "viewing" a picture in the same way each time it is shown to them, thus making them likely to reuse the same referential expression each time (Duff, Hengst, Tranel, & Cohen, 2006; Knutsen & Le Bigot, 2016; Rogers & Fay, 2016). Regardless of why it may occur, self- vs. partner-generation was coded in the current study in order to

determine whether the effect of the number of referential expressions initially considered remained significant even when initial generation was controlled for. In addition to this, in this experiment, some referential expressions might have been a better match than others to describe the Tangram figures used, potentially making the participants more like to generate and reuse these referential expression regardless of their own preferences. As explained below in the method section, this possibility was also controlled for in the current study.

Method

Participants

A total of 96 native French speakers (19 males; mean age 18.71, *SD* = 1.44) divided into 48 dyads took part in the experiment in exchange for partial course credit. They all signed an informed consent form before taking part in the experiment and were fully debriefed after the end of the experiment.

Apparatus and materials

Apparatus. The interactions between the participants were recorded using two digital voice recorders.

Materials used in the Dialogue Phase. A pool of 60 Tangram figures, chosen randomly, was used in the study. These figures were randomly divided into three sets (Sets 1, 2 and 3). Within each set, each Tangram was randomly allocated a number between one and 20. Each Tangram figure was then printed along with its number on an A6 piece of white cardboard for use during the Dialogue Phase. All pictures shown to any given pair of participants during this phase would all belong to the same set.

Materials used in the Matching Phase. Slideshows including the same Tangram figures were then prepared for use during the Matching Phase. Each trial included three

slides. The first slide featured information about the role (Director or Matcher) played by each participant. A number (1, 2 or 3) was shown on the second slide (as explained in the next section, this number was used to inform the Director of the position of the target picture on the next slide). Three Tangram figures were shown on the third slide, one of which was the target (the other two pictures were distractors). The position of the target on the slide (first, second or third position) was randomised. In critical trials (10 per participant), the Director and his or her initial partner had already been shown the target during the Dialogue Phase. In non-critical trials (20 per participant), the Director and his or her initial partner had not been shown the target during the Dialogue Phase, as they belonged to a set of pictures which had not been shown to the Director and his or her initial partner. There were two familiarisation trials and 60 test trials during the Matching Phase. The participants performed the test trials in one of two random orders. The Tangram figures used in the familiarisation trials were not used in the Dialogue Phase or in the rest of the Matching Phase.

Task and procedure

Instructions. A group of four participants was recruited for each experimental session. These participants were greeted upon arrival at the laboratory by two experimenters (hereafter Y and Z). Two participants (hereafter A and B) were asked to follow Experimenter Y and the other two participants (hereafter C and D) were asked to follow Experimenter Z. The four participants were divided into these two pairs at random. Each pair was taken to a (different) quiet experimental booth, where they sat next to each other and facing the experimenter. The experimenter then explained that the participants would take part in an experiment on how people talk about abstract pictures. They were told that the experiment would be divided into several steps, but they were not told in advance what they would do in each step.

Dialogue Phase. During this phase, the experimenter showed the participants Tangram figures one by one; the participants' task consisted in coming up together with a label to describe the picture. The participants were encouraged to take an active part in the discussion and to react to their partner's suggestions (for instance, by proposing alternative labels if they could think of any). Once the participants had reached an agreement, they could move on to the following trial. Within each group of four participants, each pair was shown a different set of pictures (e.g., A and B were shown Set 1 and C and D were shown Set 2). Which pictures were shown to which pairs was counterbalanced across groups. The pictures were shown to the participants in a random order; one of the participants (A or C) was instructed to say the number of the picture out loud before the participants started discussing the picture. The purpose of this was to help the transcribers to subsequently determine which picture was discussed by the participants in each trial.

Partner switch. At the end of the Dialogue Phase, Participants B and D were asked to follow their experimenter out of the booth. In the "Different Partner" Condition, Participant B joined Experimenter Z and Participant C, and Participant D joined Experimenter Y and Participant A. In the "Same Partner" Condition, the participant and the experimenter waited outside the booth for 30 seconds before re-entering the same booth. The purpose of this was to increase comparability across conditions. In the "Different Partner" Condition, the participants were given no information about what their new partner had done while they were performing the Dialogue Phase. This was to prevent the participants from supposing that they had previously seen the same pictures (which was actually not the case).

Matching Phase. During this phase, both participants sat next to each other, facing a computer screen. The slideshow shown on the screen was controlled by the experimenter. At the beginning of each trial, the screen featured information about the role played by each participant during the trial (the participants switched roles after each trial). The experimenter

asked the participant who would play the role of Matcher to look away from the screen. The Director was then shown a number between one and three. This number corresponded to the position of the target picture on the following slide. After two seconds, the number disappeared from the screen; a "beep" sound would then inform the Matcher that he or she could now look at the screen. Three Tangram pictures then appeared on the screen. The Director's task was to describe the picture which was in the position corresponding to the number shown previously. The Matcher's task was to identify this picture on the basis of the Director's instructions and to say the corresponding number out loud. The participants moved on to the following trial once the correct answer had been given by the Matcher. The partners performed 62 trials during this phase (two familiarisation trials followed by 60 test trials).

The fact that the participants in both pairs saw different sets of pictures during the Dialogue Phase meant that when the Director described a target picture to the Matcher in the "Different Partner" condition, this picture was already known to the Director, but not to the Matcher. This is consistent with previous studies where Directors in experimental conditions similar to the current "Different Partner" Condition interacted with new matchers who had no prior knowledge of the referents under discussion (see Brennan & Clark, 1996).

None of the two phases were limited in time; the experiment lasted less than an hour. A recap of the procedure is provided in Figures 2 ("Same Partner" Condition) and 3 ("Different Partner" Condition).

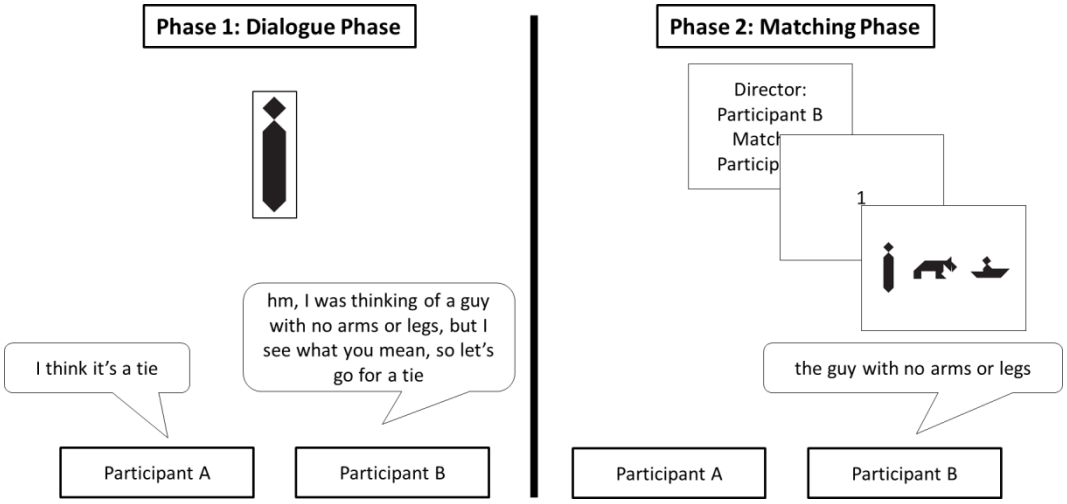

Figure 2. Recap of the procedure used in the "Same Partner" Condition.

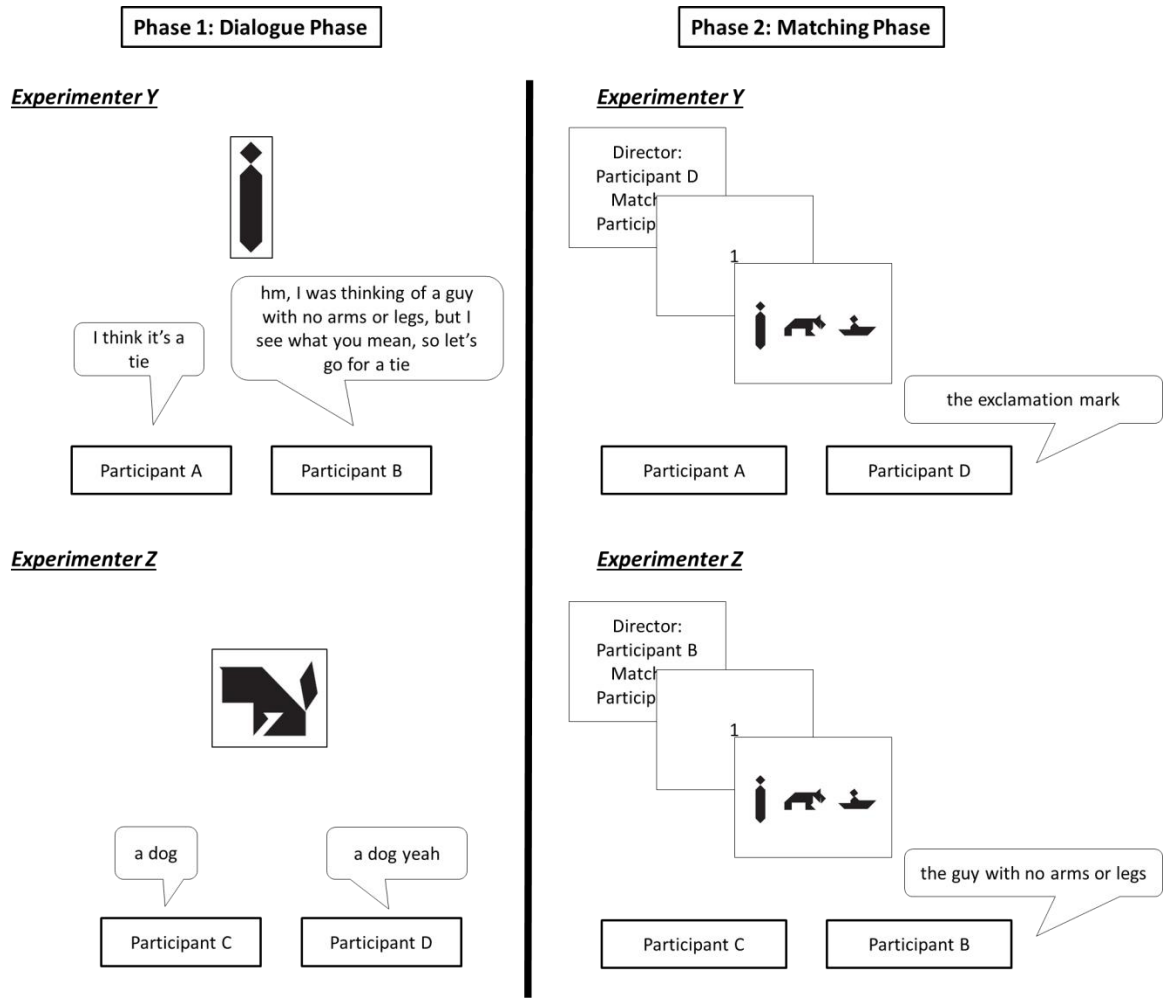

Figure 3. Recap of the procedure used in the "Different Partner" Condition.

Data coding and experimental design

Transcription and coding – Dialogue Phase. The interactions between the participants during the Dialogue Phase were transcribed and the referential expressions used were identified by a research assistant (RA) who was blind to the goal of the study and to the research hypotheses. For the purposes of this coding scheme, a referential expression was defined as a unique way of conceptualising an object (e.g., a table), person (e.g., a woman), animal (e.g., a cat) or entity (e.g., freedom) referred to by the participants (see Tolins et al., 2017, who used a similar coding scheme). For each pair of participants, the RA then coded each Tangram figure for the number of referential expressions considered. If the participants mentioned only one object, person, animal or entity, the trial was coded as a trial in which "only one referential expression was considered"; if the participants mentioned more than one object, person, animal and/or entity, the trial was coded as a trial in which "more than one referential expression was considered". For instance, in the first two examples in Tables 1 and 3, the participants considered more than one referential expression (i.e., "cat" and "duck" in the first example" and "hourglass" and "keyhole" in the second example), but they only considered one referential expression ("swan") in the third example.

Each referential expression was also coded for who had initially generated it in the dyad, which was used as a basis to code for self- vs. partner-generation during the Matching Phase (see below). For instance, in the first example provided in Tables 1 and 3, A generated the referential expression "cat" and B generated the referential expression "duck".

Moreover, based on the participants' response to the experimenter after each discussion, each referential expression was also coded for whether or not it corresponded to the pairs' preferred referential expression. For instance, in the second example provided in Tables 1 and 3, C and D finally decided that "keyhole" was their preferred referential expression for the Tangram figure under discussion. In the third example, because only one

referential expression (“swan”) was considered, this referential expression was necessarily the pair’s preferred referential expression.

Table 1

Transcript Sample – Dialogue Phase

Dyad	Condition	Set	Figure	Participant	Content of speech turn	Notes
D5	Same	S2	F17	A	17 a cat	*A produces the referential expression "cat".*
D5	Same	S2	F17	B	a duck	*B produces the referential expression "duck".*
D5	Same	S2	F17	A	a duck?	
D5	Same	S2	F17	B	where is your cat?	
D5	Same	S2	F17	A	ah yeah because you you see it on this side	
D5	Same	S2	F17	B	yeah	
D5	Same	S2	F17	A	I can see it this side here I can see the tail the legs	
D5	Same	S2	F17	B	ah yeah	
D5	Same	S2	F17	A	the ears and everything and you you see it on the other side	
D5	Same	S2	F17	B	ah yeah the cat is more visible in fact	
D5	Same	S2	F17	A	now you see it more you see it?	
D5	Same	S2	F17	B	yeah	
D5	Same	S2	F17	A	so let's say a cat	
D5	Same	S2	F17	B	a cat	
D5	Same	S2	F17	A	a cat	*"Cat" is chosen as the preferred referential expression.*
D8	Same	S3	F12	C	erm number 12	
D8	Same	S3	F12	D	erm I don't know it reminds me of an hourglass but upside down	*D produces the referential expression "upside down hourglass".*
D8	Same	S3	F12	C	erm it reminds me of a keyhole	*C produces the referential expression "keyhole".*
D8	Same	S3	F12	D	yeah me too well keyhole I think I would say keyhole yeah	
D8	Same	S3	F12	C	I agree	*"Keyhole" is chosen as the preferred*

						referential expression.
D11	Different	S2	F17	D	picture number 17	
D11	Different	S2	F17	C	it reminds me of a swan	*C produces the referential expression "swan".*
D11	Different	S2	F17	D	yeah me too the swan	
D11	Different	S2	F17	C	we agree on the name	
D11	Different	S2	F17	D	it's okay	
D11	Different	S2	F17	C	me too	*"Swan" is chosen as the preferred referential expression.*

Note. All examples are translated from French.

Table 2

Transcript Sample – Matching Phase

Dyad	Condition	Set	Figure	Participant	Role	Content of speech turn	Notes
D5	Same	2	17	A	Director	it looks like a boat	*A produces the referential expression "boat".*
D5	Same	2	17	B	Matcher	number 2	
D8	Same	3	12	D	Director	erm it looks like a keyhole	*D produces the referential expression "keyhole".*
D8	Same	3	12	C	Matcher	the third one	
D11	Different	2	17	C	Director	it reminds me of a swan	*C produces the referential expression "swan".*
D11	Different	2	17	B	Matcher	number 3	

Note. All examples are translated from French.

Table 3

Coding Example – Dialogue and Matching Phase

Dyad	Set	Figure	Referential expression	Number of referential expressions	Who generated the referential expression?	Preferred referential expression?	Frequency in the corpus	Director in Matching Phase	Reused during the Matching Phase?
D5	S2	F17	Cat	More than one	A	Yes	12	A	No
D5	S2	F17	Duck	More than one	B	No	7	A	No
D8	S3	F12	Upside down hourglass	More than one	D	No	1	D	No
D8	S3	F12	Keyhole	More than one	C	Yes	3	D	Yes
D11	S2	F17	Swan	One	C	Yes	15	C	Yes

Note. All examples are translated from French.

What is more, as mentioned in the introduction section, some referential expressions might have been a better match than others to describe the Tangram figures used in the current studies. For instance, most people might think that "a tie" is a more relevant description of one of the Tangram figures used than "a man with no arms or legs". The participants might have been more likely to agree upon the most relevant referential expression during the Dialogue Phase (it might also be that some Tangram figures used were more "obvious" than others, thus making the participants less likely to come up with more than one referential expression when discussing these); they might then have reused the preferred referential expression during the Matching Phase not because it was their preferred referential expression, but simply because it happened to be the most relevant referential expression to describe that figure in particular. In order to discard this possibility, the RA listed all the referential expressions used during the Dialogue Phase to describe all the Tangram figures used in this study, and counted how many times each referential expression was used to describe each Tangram figure (e.g., the RA counted how many times pairs of participants came up with the referential expressions "a tie" and "a man with no arms or legs" to describe the Tangram Figure mentioned previously). This count, or *frequency in the corpus*, provided an estimate of how suitable a referential expression was perceived for each Tangram figure in the sample of participants used. The counterbalancing used meant that each Tangram figure was shown to 16 pairs in total (16 pairs x 3 sets of figures = 48 pairs in total), implying that any referential expression listed may have been considered by a minimum of one pair and by a maximum of 16 pairs. This variable was introduced in the analysis in order to determine whether the effect of the number of referential expressions initially considered remained significant even when frequency in the corpus was controlled for. An example is provided in Table 4.

Table 4

List of Referential Expressions Considered by the Participants to Refer to one of the Tangram Figures and Frequency in the Corpus

Tangram figure	Referential expression	Frequency in the corpus
 *Set 3, Figure 14*	House	9
	Helmet	5
	Door	2
	Temple	2
	Circus	1
	Chimney	1
	Oven	1
	Goldorak (cartoon character)	1
	Keyhole	1
	Tunnel	1

Note. All examples are translated from French. Frequencies in the corpus add up to more than 16 (which was the number of pairs each picture was shown to) because each pair had the opportunity to produce more than one referential expression to describe each picture.

Transcription and coding – Matching Phase. The interactions between the participants during the Matching Phase were also transcribed. Each critical trial was then coded for whether or not the Director reused a referential expression considered during the Dialogue Phase. Only the referential expressions produced by the Director were taken into account here, as the Matcher seldom referred to the target figure during this phase. A sample transcript is provided in Tables 2 and 3 to illustrate the coding scheme used in this phase. In the first example, A reused neither of the two referential expressions which had been considered during the Dialogue Phase. In the second example, D reused the referential expression “keyhole”, which was initially generated by C, and which was the pair’s preferred referential expression for this Tangram figure. Finally, in the third example, C reused the referential expression “swan”, which he or she initially generated him- or herself, and which was also the pair’s preferred referential expression for this Tangram figure.

The data from 12 dyads (6 in the "Same Partner" Condition and 6 in the "Different Partner" Condition; this represented 25% of the entire corpus) were double-coded by the RA and the first author. Cohen's kappa was then calculated in order to determine to compare the two sets of coding. The value obtained was 0.82, reflecting almost perfect agreement. All disagreements were discussed and resolved by the two coders. The remaining data were single-coded by the RA.

Experimental design and dependent variables. Four IVs were used in the main analysis. The first one was the experimental condition. This was a categorical IV with two levels: participants either performed the experiment in the "Same Partner" Condition or in the "Different Partner" Condition. The second IV was whether or not more than one referential expression was considered during the Dialogue Phase. This was a categorical IV with two levels: one referential expression considered vs. more than one referential expression considered. The third IV was the identity of the participant who initially generated the referential expression during the Dialogue Phase. This was a categorical IV with two levels: from each participant's point of view, each referential expression was either self-generated or partner-generated. Finally, the fourth IV was referential expression frequency in the corpus. This was a continuous IV, which was standardised for the purpose of the analysis. The binary DV was whether or not the Director reused the preferred referential expression during the Matching Phase.

Results

Descriptive statistics

The data from 45 figures (4.69% of the entire dataset) were discarded from the analysis due to recording issues or experimenter error during the Dialogue Phase and/or during the

Matching Phase, making it difficult to identify the referential expressions produced. All figures reported hereafter were calculated after these data were removed.

The total number of words produced by the participants during the Dialogue Phase was 38,034 (19,348 in the "Same Partner" Condition and 18,686 in the "Different Partner" Condition). The average number of words produced by dyad was 792.38 (*SD* = 332.55; in the "Same Partner" Condition, the average was 806.17 [*SD* = 331.79]; in the "Different Partner" Condition, the average was 778.58 [*SD* = 339.86]).

Altogether, the participants described 915 Tangram figures during the Dialogue Phase (48 pairs of participants x 20 figures described per pair – 45 figures which were removed from the dataset). The average number of words produced to describe each Tangram figure was 39.62 (*SD* = 34.96; in the "Same Partner" Condition, the average was 40.31 [*SD* = 35.64]; in the "Different Partner" Condition, the average was 38.93 [*SD* = 38.93]). The participants considered one referential expression only for 540 Tangram figures (273 in the "Same Partner" Condition and 267 in the "Different Partner" Condition) and more than one referential expression for 375 Tangram figures (189 in the "Same Partner" Condition and 186 in the "Different Partner" Condition).

The total number of words produced during the Matching Phase (including both critical and non-critical trials) was 39,445 (18,717 in the "Same Partner" Condition and 20,728 in the "Different Partner" Condition). The average number of words produced by dyad was 821.77 (*SD* = 374.98; in the "Same Partner" Condition, the average was 779.88 [*SD* = 280.55]; in the "Different Partner" Condition, the average was 863.67 [*SD* = 452.72]). The average number of words produced to describe each Tangram figure was 13.70 (*SD* = 10.05; in the "Same Partner" Condition, the average was 13.00 [*SD* = 8.43]; in the "Different Partner" Condition, the average was 14.36 [*SD* = 11.41]).

Preliminary analysis: Effects of common ground on the Matching Task

Three analyses were conducted to check that common ground was built during the Dialogue Phase and exploited during the Matching Phase. These three analyses, which are reported in Appendix A, provide evidence that the corpus gathered in the current study exhibited features similar to those found in other studies on dialogic common ground (e.g., Clark & Wilkes-Gibbs, 1986; Hupet & Chantraine, 1992).

Overview of the main analysis

Mixed models were used to conduct the analyses in this study. The statistical software used was SAS 9.4 (GLIMMIX procedure). One of the main advantages of mixed models is that they account for potential variability across analysis units through the inclusion of random intercepts and for the fact that analysis units potentially differ in their sensitivity to IVs through the inclusion of random slopes. Following Barr, Levy, Scheepers, and Tily's (2013) recommendations, an initial model including the maximal random effects structure justified by the design was used to analyse the data. These multilevel models involved three units of analysis (dyads, participants – which were nested within dyads – and items – which were the Tangram pictures used in the experiment); by-dyad, by-participant and by-item random intercepts as well as by-dyad, by-participant and by-item random slopes corresponding to within-units IVs were therefore included in the model. However, using such maximal random effects structure often causes mixed models to fail to converge. Convergence failure reflects the software's inability to find an appropriate model to fit the data within the specified number of iterations. This is often due to the variance associated with at least one of the random effects (random intercepts or slopes) being equal to zero (Kiernan, Tao, & Gibbs 2012). When convergence problems arose in the current analysis, these effects were removed (they are automatically identified in SAS) and the analysis was conducted again. Removing

problematic random effects does not affect the model parameters; that is, even if the degrees of freedom of the model are inflated when these random effects are removed, the actual parameters of the model remain unchanged. For transparency reasons, we nonetheless report the full random structure model output in Appendix B, even though this model failed to converge.

Logistic mixed models were used in the analysis, as the DV used was binary (Jaeger, 2008). One of the indicators returned by logistic models is the odds ratio (OR), which quantifies the odds of one event occurring rather than another. For instance, obtaining an odds ratio of 2.00 when examining the effect or whether or not more than one referential expression was initially considered would mean that the preferred referential expression was twice as likely to be reused when it was the only referential expression to have been considered initially. ORs are also an indicator of effect size: the larger the OR, the larger the corresponding effect (Agresti, 2002).

Only the data from the 20 critical trials were analysed, as the hypothesis solely focused on whether or not Directors reused the referential expressions considered previously in these trials; the data from the remaining 40 non-critical trials were removed from the dataset. What is more, only the data corresponding to the preferred referential expressions were analysed, as the hypothesis focused mainly on the likelihood of reusing these; the analysis of the reuse of dispreferred referential expressions is nonetheless reported in Appendix C for information purposes only.

The experimental design used in the experiment was unbalanced (this is because there were fewer figures for which the participants initially considered more than one referential expression than figures for which the participants initially considered one referential expression only); to account for this, degrees of freedom were corrected using Satterthwaite's correction (Keselman, Algina, Kowalchuk, & Wolfinger, 1999; Satterthwaite, 1946).

The main effects corresponding to the four IVs were included in the analysis; interactions were only included in the analysis if they were statistically significant.

Results of the main analysis

The model included Condition, the number of referential expressions initially considered, the identity of the participant who initially generated the referential expression, the referential expression's frequency in the corpus and the Condition x frequency in the corpus interaction as fixed effects. The outcome variable was the probability of the Director reusing the preferred referential expression during the Matching Phase. The random effects structure included by-participant and by-item random intercepts and by-item random slopes corresponding to Condition, the number of referential expressions initially considered and the referential expression's frequency in the corpus. The data are shown in Table 5. The model parameters are shown in Table 6.

Table 5

Proportion of Matching Phase Trials where the Preferred Referential expression was Reused as a Function of the Number of Referential expressions Initially Considered, of the Condition and of Who Initially Generated the Referential Expression

	Same Partner			Different partner			**Total**
	Self-gen.	Partner-gen.	**Total (same)**	Self-gen.	Partner-gen.	**Total (different)**	
One referential expression considered	0.82	0.73	**0.77**	0.81	0.67	**0.75**	**0.76**
More than one referential expression considered	0.62	0.56	**0.59**	0.59	0.44	**0.51**	**0.55**
Total	**0.73**	**0.67**	**0.70**	**0.73**	**0.58**	**0.65**	**0.67**

Table 6

Model Parameters

Effect	Estimate	Standard error	*p*
Intercept	0.20	0.20	.320
Condition: Different Partner	-0.23	0.21	.276
Condition: Same Partner	0		
Number of referential expressions considered: One	0.53	0.18	.005
Number of referential expressions considered: More than one	0		
Generation: Self	0.66	0.17	< .001
Generation: Other	0		
Frequency in the corpus	0.60	0.14	< .001
Frequency in the corpus x Condition: Different Partner	0.41	0.18	.023
Frequency in the corpus x Condition: Same Partner	0		

Firstly, a significant effect of whether or not more than one referential expression was considered during the Dialogue Phase was found, $F(1, 56) = 8.76$, $p = .005$. As predicted, Directors were more likely to reuse the preferred referential expression to describe a target figure when only one referential expression was initially considered than when several referential expressions were initially considered, $OR = 1.70$, $CI_{.95} = 1.19, 2.43$.

Secondly, a significant generation effect was found, $F(1, 909) = 15.78$, $p < .001$. As predicted, Directors were more likely to reuse self-generated referential expressions than partner-generated ones, $OR = 1.94$, $CI_{.95} = 1.40, 2.69$.

Thirdly, a significant effect of frequency in the corpus was found, $F(1, 60) = 50.55$, $p < .001$. As predicted, the likelihood of Directors reusing the preferred referential expression increased as the referential expression's frequency in the corpus also increased, $b = 0.60$.

Fourthly, the main effect of Condition failed to reach statistical significance, $F(1, 60) = 1.21$, $p = .276$.

Finally, a significant Condition x frequency in the corpus was found, $F(1, 495) = 5.20$, $p = .023$. An inspection of the *b* coefficient revealed that the frequency effect was stronger in the "Different Partner" Condition than in the "Same Partner" Condition, $b = 0.41$.

Discussion

When two people engage in an interaction, they must make decisions about how to refer to things. Initial referential decisions are made collaboratively, with both partners actively contributing to the referring process (Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). Crucially for the current study, the dialogue partners may or may not consider more than one referential expression during this process. The purpose of this work was to examine how this might affect subsequent decisions to reuse the referential expressions considered initially. The first hypothesis was that associations in memory between a referent and the preferred

referential expression is weaker when several different references were initially discussed. The results corroborated this hypothesis. In line with the memory-based approach to dialogue (Barr et al., 2014; Gorman et al., 2013; Horton, 2008; Horton & Gerrig, 2005a, 2005b, 2016; Horton & Slaten, 2012; Knutsen & Le Bigot, 2016), these results contribute to a better understanding of how low-level memory processes influence referential decisions. Building on recent studies which have shed light on the factors which contribute to reinforcing reference-partner associations in memory (Barr et al., 2014; Gorman et al., 2013), the current study is one of the first to have examined a factor contributing to the strength of memory associations between specific referential expressions and specific *referents*. These results are also in line with the idea that collaboration is an essential determinant of referential decisions (Clark, 1996; Clark & Wilkes-Gibbs, 1986). It is noteworthy that most research on presentation and acceptance seems to suggest that considering one referential expression only or considering more than one referential expression at the time of initial contribution leads to the same result – that is, the preferred referential expression is added to the partners' common ground (see Examples 1-3 in the Introduction section). The current research challenges this suggestion by showing that the number of referential expressions considered during the initial contribution process has a direct influence on their subsequent accessibility in memory, in line with the idea that dialogue partners hold representations of their common ground in which different pieces of information are more or less readily accessible (Brown-Schmidt, 2012; Knutsen et al., 2016).

Other factors than the number of referring expressions initially considered were also found to affect reuse in the current experiment. Firstly, a generation effect was found (Slamecka & Graf, 1978), in line with previous research showing that the identity of the speaker who initially generated a referring expression has a direct impact on its subsequent reuse in dialogue (e.g., Knutsen & Le Bigot, 2016; Knutsen et al., 2016). Secondly, the more

often a referential expression was considered by all dyads during the Dialogue Phase, the more likely it was to be reused during the Matching Phase, regardless of whether or not it corresponded to the partners' preferred referring expression. We have suggested that this *frequency in the corpus* variable was a good indicator of how relevant each referential expression was perceived to be to describe a specific Tangram figure by the group of participants used in the experiment. Although this variable was not of prime interest in the current study (it was initially introduced to control for the perceived relevance of referential expression), this finding is nonetheless theoretically interesting. Indeed, it implies that most participants knew what the best referential expression for each Tangram figure was, even if this was not the referential expression they had agreed upon with their partner during the Dialogue Phase. They were then capable of resorting to this knowledge to improve communication during the Matching Phase. They might have done this incidentally: the most relevant referring expression might simply have reflected the way in which most participants "viewed", or conceptualised, a Tangram figure when they saw it for the first time; when referring to the same figure again, the participants would have conceptualised it in the same way, thus making them more likely to favour the production of the most relevant referential expression. It is noteworthy that in this kind of situation, the most relevant referential expression might not only be the easiest referential expression to *produce* (from the speaker's point of view), but also the easiest referential expression to *understand* (from the listener's point of view). Thus, the relevance effect found in this study would incidentally lead the speaker to favour the production of a referential expression which can be understood easily, even in the absence of an explicitly negotiated common ground (for a discussion of incidental adaptive behaviours in dialogue, see Horton & Gerrig, 2005b; Pickering & Garrod, 2004; Shintel & Keysar, 2009) The participants might also have used their knowledge about referential expression relevance purposefully: they might simply have assumed that the most

relevant referential expression would be easy to understand for their current dialogue partner. In any event, this effect being stronger in the "Different Partner" Condition than in the "Same Partner" Condition suggests that speakers mainly rely on this kind of cue to guide the production of referential expressions when they share no linguistic common ground with their current dialogue partner. In sum, the fact that the effect of the number of referential expressions initially considered remained statistically significant when generation and relevance were controlled for confirms that this effect cannot be attributed solely to these two factors.

What is more, the identity of the person in the presence of whom the participants had the opportunity to reuse the referential expressions considered previously was also manipulated. The second main hypothesis was that the effect of the number of referential expressions initially considered would be attenuated when the dialogue partner remained the same throughout the interaction. Indeed, referential expressions are associated in memory not only with specific referents, but also with specific dialogue partners (Horton, 2007; Horton & Gerrig, 2005b), so the presence of a dialogue partner might be sufficient to compensate for a weak referential expression – referent association in memory. However, the data reported here offered no support for this hypothesis: the results suggest that the effect of the number of referential expressions initially considered can be observed in the presence of the same partner, or in the presence of a different partner. This bridges further the gap between the collaborative and the memory-based approaches to dialogue (Clark, 1996; Clark & Wilkes-Gibbs, 1986; Horton, 2008; Horton & Gerrig, 2016), by suggesting that the effect of presentation and acceptance on subsequent reuse is not partner-specific (for examples of partner-specific referential communication, see Brennan & Hanna, 2009; Brown-Schmidt, 2009; Metzing & Brennan, 2003). Rather, collaboration with one partner affects subsequent referential decisions – even when reuse occurs in the presence of a different partner. In other

words, references are negotiated collaboratively within pairs of speakers, but the consequences of such collaboration may extend to subsequent interactions with new partners, due to low-level, ordinary memory functioning.

The findings reported here raise a number of new theoretical questions, including the question of *why* initially considering several referential expressions would result in weaker referent – referential expression associations in memory. A first possible explanation is that dialogue partners are less certain of which referential expression was finally agreed upon when several different ones were initially considered, preventing them from building a clear memory representation in which a specific referent is associated with one specific referential expression (i.e., the preferred referential expression). However, whilst this might be the case in real-life dialogue, where dialogue partners might not explicitly state which referential expressions(s) they agree upon, it is unlikely that such uncertainty drove the results reported in the current experiment, as the participants were explicitly required to specify to the experimenter which referential expression they had agreed upon at the end of each trial during the Dialogue Phase. Another, more plausible explanation is related to research on the *fan effect* (e.g., Anderson & Reder, 1999; Bunting, Conway, & Heitz, 2004; Radvansky & Zacks, 1991). This research has shown that the more facts people learn about a particular concept, the longer it takes them to retrieve any given fact about this concept from their memory; this also results in an increase in the number of retrieval mistakes made. Although the current study involved no measure of retrieval time, a similar mechanism may be at play in dialogue whereby associating several referential expressions with a single referent might cause the preferred referential expression to become more difficult and/or to take more time to retrieve for dialogue partners. This explanation would have two important theoretical implications. Firstly, it would imply that just like resonance (Ratcliff, 1978) and the production/generation effect (MacLeod, 2011; Slamecka & Graf, 1978), another *ordinary*

memory mechanism – namely the fan effect – has an impact on the referential decisions made during an interaction. Secondly, it would imply that dispreferred referential expressions, just like preferred ones, are encoded in the partners' conversational memory, insofar as their presence there has an impact on subsequent referential expression reuse. This second point could be of particular relevance for the collaborative approach to dialogue (Clark, 1996; Clark & Wilkes-Gibbs, 1986), in which the question of whether or not dispreferred referential expressions are deemed part of the partners' common ground remains open. It is important to highlight that the content of each speaker's conversational memory is *not* necessarily part of the partners' common ground, as the latter is usually defined as the knowledge that two partners share and are aware of sharing (i.e., a piece of information might be available in one of the partners' conversational memory, but not in the other partner's conversational memory; moreover, a piece of information might be available in both partners' memory, but they might not be aware that they share it). It is nonetheless noteworthy that dispreferred referential expressions, just like preferred referential expressions, may affect the remainder of the interaction. Future research will seek to determine whether there is a direct link between the *number* of referential expressions considered and the likelihood that the preferred referential expression is reused subsequently (this link could not be examined in the current study, which only distinguished between two kinds of situations: situations in which only one referential expression is considered and situations in which more than one referential expression is considered). Such a link would provide additional strong evidence in support of the fan effect hypothesis.

Importantly, the current study also presents a number of limitations. Among these was the fact that the participants were explicitly required to agree upon a single referential expression to describe the pictures discussed during the Dialogue Phase. In more naturalistic dialogue settings, the contribution process would be much more spontaneous and would

probably be characterized by a number of features which could not be observed in the current laboratory study. For instance, each dialogue partner might have a particular interest in making sure that their referential expression is finally chosen, which would affect both the contribution process and subsequent reuse.

To conclude, this study sought to examine separately two situations which are not usually distinguished in the dialogue literature: situations in which only one referential expression is considered during the contribution process on one hand and situations in which several different referential expressions are considered before an agreement can be reached on the other hand. The results of the experiment reported here show that this affects subsequent referential decisions, due to difference in reference accessibility in memory. Although these results are in part compatible with a collaborative view of human dialogue (Clark, 1996), they also advocate for the idea that low-level memory processes incidentally contribute to dialogue success (Horton & Gerrig, 2016).

Acknowledgments

We thank Jessica Soira and Anaïs Bourgoin for their help with data collection and transcription and Sandrine Belkadi for her assistance with coding. We also thank two anonymous reviewers for their very helpful comments on a previous version of the manuscript.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed). New York, NY: John Wiley & Sons.

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*, 186–197.

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, *27*, 195–225. https://doi.org/10.1207/s15516709cog2702_3

Bangerter, A., Clark, H. H., & Katz, A. R. (2004). Navigating joint projects in telephone conversations. *Discourse Processes*, *37*, 1–23. https://doi.org/10.1207/s15326950dp3701_1

Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental: General*, *143*, 404–413. https://doi.org/10.1037/a0031813

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

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*, 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Brennan, S. E., & Ohaeri, J. O. (1999). Why do electronic conversations seem less polite? The costs and benefits of hedging. In *Proceedings of the International Joint Conference on Work Activities Coordination and Collaboration* (pp. 227–235). New York, NY: ACM. https://doi.org/10.1145/295665.295942

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language, 61*, 171–190. https://doi.org/10.1016/j.jml.2009.04.003

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes, 27*, 62–89. https://doi.org/10.1080/01690965.2010.543363

Bunting, M. F., Conway, A. R. A., & Heitz, R. P. (2004). Individual differences in the fan effect and working memory capacity. *Journal of Memory and Language, 51*, 604–622. https://doi.org/10.1016/j.jml.2004.07.007

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: University of Chicago Press.

Clark, H. H. (1996). *Using language*. Cambridge, MA: Cambridge University Press.

Clark, H. H., & Bangerter, A. (2004). Changing conceptions of reference. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 25–49). Basingstoke, England: Palgrave Macmillan.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62–81. https://doi.org/0.1016/j.jml.2003.08.004

Clark, H. H., & Marshall, C. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (Vol. 2, pp. 10–63). Cambridge, MA: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259–294. https://doi.org/10.1207/s15516709cog1302_7

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience*, *9*, 140–146. https://doi.org/10.1038/nn1601

Fox Tree, J. E. (2010). Discourse markers across speakers and settings. *Language and Linguistics Compass*, *3*, 1–13. https://doi.org/10.1111/j.1749-818X.2010.00195.x

Gorman, K. S., Gegg-Harrison, W., Marsh, C. R., & Tanenhaus, M. K. (2013). What's learned together stays together: Speakers' choice of referring expression reflects shared experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 843–853. https://doi.org/10.1037/a0029467

Horton, W. S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes*, *22*, 1114–1139.

Horton, W. S. (2008). A memory-based approach to common ground and audience design. In I. Kecskes (Ed.), *Intention, common ground, and the egocentric speaker-hearer* (pp. 189–222). Berlin, Germany/New York, NY: Mouton de Gruyter.

Horton, W. S., & Gerrig, R. J. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes*, *40*, 1–35. https://doi.org/10.1207/s15326950dp4001_1

Horton, W. S., & Gerrig, R. J. (2005b). The impact of memory demands on audience design during language production. *Cognition*, *96*, 127–142. https://doi.org/10.1016/j.cognition.2004.07.001

Horton, W. S., & Gerrig, R. J. (2016). Revisiting the memory-based processing approach to common ground. *Topics in Cognitive Science*, *8*, 780–795. https://doi.org/10.1111/tops.12216

Horton, W. S., & Slaten, D. G. (2012). Anticipating who will say what: The influence of speaker-specific memory associations on reference resolution. *Memory & Cognition*, *40*, 113–126. https://doi.org/10.3758/s13421-011-0135-7

Hupet, M., & Chantraine, Y. (1992). Changes in repeated references: Collaboration or repetition effects? *Journal of Psycholinguistic Research*, *21*, 485–496. https://doi.org/10.1007/BF01067526

Hupet, M., Chantraine, Y., & Nef, F. (1993). References in conversation between young and old normal adults. *Psychology and Aging*, *8*, 339–346.

Hupet, M., Seron, X., & Chantraine, Y. (1991). The effects of the codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology*, *82*, 449–462. https://doi.org/10.1111/j.2044-8295.1991.tb02412.x

Isaacs, E. A., & Clark, H. H. (1987). Reference in conversations between experts and novices. *Journal of Experimental Psychology: General*, *116*, 26–37. https://doi.org/10.1037/0096-3445.116.1.26

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1999). The analysis of repeated measurements: A comparison of mixed-model satterthwaite F tests and a nonpooled adjusted degrees of freedom multivariate test. *Communications in*

Statistics - Theory and Methods, *28*, 2967–2999. https://doi.org/10.1080/03610929908832460

Kiernan, K., Tao, J., & Gibbs, P. (2012). Tips and strategies for mixed modelling with SAS/STAT procedures. Presented at the 2012 SAS Global Forum, Orlando, FL.

Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin & Review*, *21*, 1590–1599. https://doi.org/10.3758/s13423-014-0620-7

Knutsen, D., & Le Bigot, L. (2016). Conceptual match as a determinant of reference reuse in dialogue. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000313

Knutsen, D., Ros, C., & Le Bigot, L. (2016). Generating references in naturalistic face-to-face and phone mediated dialogue settings. *Topics in Cognitive Science*, *8*, 796–818. https://doi.org/10.1111/tops.12218

Liu, K., & Fox Tree, J. E. (2012). Hedges enhance memory but inhibit retelling. *Psychonomic Bulletin & Review*, *19*, 892–898. https://doi.org/10.3758/s13423-012-0275-1

MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review*, *18*, 1197–1202. https://doi.org/10.3758/s13423-011-0168-8

McInnes, F., & Attwater, D. (2004). Turn-taking and grounding in spoken telephone number transfers. *Speech Communication*, *43*, 205–223. https://doi.org/10.1016/j.specom.2004.04.001

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*, 201–213. https://doi.org/10.1016/S0749-596X(03)00028-7

Mills, G. J. (2014). Dialogue in joint activity: Complementarity, convergence and conventionalization. *New Ideas in Psychology*, *32*, 158–173. https://doi.org/10.1016/j.newideapsych.2013.03.006

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–190. https://doi.org/10.1017/S0140525X04000056

Radvansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 940–953. https://doi.org/10.1037/0278-7393.17.5.940

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Rogers, S. L., & Fay, N. (2016). Stick or switch: A selection heuristic predicts when people take the perspective of others or communicate egocentrically. *PLOS ONE*, *11*, e0159570. https://doi.org/10.1371/journal.pone.0159570

Rossnagel, C. (2000). Cognitive load and perspective-taking: Applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, *30*, 429–445. https://doi.org/10.1002/(SICI)1099-0992(200005/06)30:3<429::AID-EJSP3>3.0.CO;2-V

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, *2*, 110–114. https://doi.org/10.2307/3002019

Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science*, *1*, 260–273. https://doi.org/10.1111/j.1756-8765.2009.01018.x

Slamecka, N., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Tolins, J., Zeamer, C., & Fox Tree, J. E. (2017). Overhearing dialogues and monologues: How does entrainment lead to more comprehensible referring expressions? *Discourse Processes*. https://doi.org/10.1080/0163853X.2017.1279516

Tylén, K., Weed, E., Wallentin, M., Roepstorff, A., & Frith, C. D. (2010). Language as a tool for interacting minds. *Mind and Language*, *25*, 3–29. https://doi.org/10.1111/j.1468-0017.2009.01379.x

Ulrich, R., & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology*, *37*, 513–525. https://doi.org/10.1006/jmps.1993.1032

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*, 183–194. https://doi.org/10.1016/0749-596X(92)90010-U

Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 919–937. https://doi.org/10.1037/a0036161