

Appendix A: Preliminary analysis

Three analyses were conducted to check that common ground was built during the Dialogue Phase. Past research has shown that Directors involved in matching tasks produce fewer words to describe pictures when they share common ground with their current Matcher (e.g., Clark & Wilkes-Gibbs, 1986; Hupet & Chantraine, 1992). Accordingly, the preliminary analyses compared the features of the Directors' speech in critical trials (in which the Director described a picture he or she had already discussed during the Dialogue Phase) and in non-critical trials (in which the Director described a picture that was new to him or her) during the Matching Phase. The total number of words produced by the Director per trial was used as the DV in the first preliminary analysis. What is more, in this kind of task, the Directors' descriptions include not only content words, but also other words such as fillers (e.g., "erm") or hedges (i.e., words and phrases used by dialogue partners to specify that an utterance is provisional, such as "a kind of"; Brennan & Ohaeri, 1999; Liu & Fox Tree, 2012). The purpose of the second analysis was to verify that common ground led Directors to use fewer content words, which form the gist of the description, in critical trials. In order to do this, the Directors' speech was coded for content words. This category included proper nouns (e.g., "Paris"), common nouns (e.g., "tie"), adjectives (e.g., "small") and verbs (e.g., "to eat"). Finally, the purpose of the third analysis was to verify that common ground led Directors to use fewer hedges in critical trials. In order to do this, the Directors' speech was coded for hedges. This category included words and expressions such as "a kind of", "it could be described as", "I think", "a sort of" or "like".

Common ground was expected to affect the Directors' speech mainly in the "Same Partner" Condition, but not in the "Different Partner" Condition, where target pictures did not belong to the common ground shared by the Director and the Matcher.

The data were analysed using mixed models and following the same rationale as the main analysis. The only difference was that the models used in the preliminary analysis were linear mixed models (and not logistic mixed models, as in the main analysis), because the DVs used were continuous rather than binary.

Effect of common ground on the number of words produced by the Director

The average number of words produced by the Director during the Matching Phase in critical and non-critical trials is reported in Table A1. A preliminary inspection of the data suggested that a large majority of descriptions (90%) included between one and 20 words, but some descriptions included up to 90 words. Although mixed models are less sensitive to departures from normality than standard analysis methods (ANOVA, regression), a lognormal function, which is skewed to the right, was used to account for departure from normality (see Ulrich & Miller, 1993).

Table A1

Average Number of Words Produced by the Directors per Trial during the Matching Phase as a function of Trial Type and Condition

	“Same Partner”	“Different Partner”	Total
Critical trials	8.98 (6.84)	10.36 (9.29)	9.64 (8.14)
Non-critical trials	11.29 (7.66)	12.43 (11.03)	11.87 (9.54)
Total	10.52 (7.48)	11.80 (10.57)	11.16 (9.18)

Note. Standard deviations are reported in brackets.

The model included Trial type and Condition as fixed effects. The outcome variable was the number of words produced by the Directors in each trial of the Matching Phase. The random effects structure included by-dyad random intercepts, by-dyad, by-participant and by-item random slopes corresponding to Trial Type and by-item random slopes corresponding to

Condition. A main effect of Trial Type was found, $F(1, 40) = 48.66, p < .001$. Directors produced more words in non-critical trials than in critical trials, $b = 0.20$. The effect of Condition failed to reach statistical significance, $F(1, 46) = 0.33, p = .571$.

Effect of common ground on the number of content words produced by the Director

The average number of content words produced by the Director during the Matching Phase in critical and non-critical trials is reported in Table A2. A preliminary inspection of the data revealed that although the variability in this dataset was smaller than in the first analysis, the data were still not distributed normally. This was because although a large majority of descriptions (90%) included between zero and four content words, some descriptions included up to 13 content words. Following the same rationale as in the first analysis, a lognormal function was thus used in this analysis.

Table A2

Average Number of Content Words Produced by the Directors per Trial during the Matching Phase

	“Same Partner”	“Different Partner”	Total
Critical trials	2.21 (1.53)	2.38 (1.57)	2.29 (1.55)
Non-critical trials	2.49 (1.70)	2.54 (1.73)	2.51 (1.72)
Total	2.40 (1.65)	2.49 (1.68)	2.44 (1.67)

Note. Standard deviations are reported in brackets.

The model included Trial Type, Condition and the interaction between these two factors as fixed effects. The outcome variable was the number of content words produced by the Director in each trial of the Matching Phase. The random effects structure included by-dyad, by-participant and by-item random intercepts, by-dyad and by-participant random slopes corresponding to Trial Type and by-item random slopes corresponding to Condition. A

significant effect of Trial Type was found, $F(1, 42) = 14.49, p < .001$. Directors produced more content words in non-critical trials than in critical trials. The main effect of Condition failed to reach statistical significance, $F(1, 46) < 0.01, p = .951$. However, there was a significant Trial Type x Condition interaction, $F(1, 42) = 4.34, p = .043$ (Table A2). An inspection of the b coefficient revealed that the difference between critical and non-critical trials was smaller in the “Different Partner” Condition than in the “Same Partner” Condition.

Effect of common ground on the number of hedges produced by the Director

The average number of hedges produced by the Director during the Matching Phase in critical and non-critical trials is reported in Table A3. A preliminary inspection of the data revealed that although a large majority of descriptions (93%) included either zero or one hedge, some descriptions included up to 10 hedges. Following the same rationale as in the previous two analyses, a lognormal function was thus used in this analysis.

Table A3

Average Number of Hedges Produced by the Directors per Trial during the Matching Phase

	“Same Partner”	“Different Partner”	Total
Critical trials	0.48 (0.60)	0.44 (0.65)	0.46 (0.63)
Non-critical trials	0.61 (0.69)	0.61 (0.88)	0.61 (0.79)
Total	0.57 (0.66)	0.56 (0.82)	0.56 (0.75)

Note. Standard deviations are reported in brackets.

The model included Trial Type and Condition as fixed effects. The outcome variable was the number of hedges produced by the Directors in each trial of the Matching Phase. The random effects structure included by-dyad, by-participant and by-item random intercepts and by-participant random slopes corresponding to Trial Type. A main effect of Trial Type was found, $F(1, 72) = 4.64, p = .035$. Directors produced more words in non-critical trials than in

critical trials, $b = 0.05$. The effect of Condition failed to reach statistical significance, $F(1, 45) = 2.28, p = .138$.

Discussion of the results

These preliminary analyses confirmed that the common ground built during the Dialogue Phase affected the Directors' behaviour during the Matching Phase. Indeed, Directors reused fewer content words during the Matching Phase when they described pictures which they had already discussed during the Dialogue Phase. What is more, this was mainly the case in the "Same Partner" Condition, where the Director and Matcher's common ground included referential expressions corresponding to these pictures. This is consistent with the finding that Directors produce more concise descriptions when they share common ground with Matchers in this kind of task (e.g., Clark & Wilkes-Gibbs, 1986; Hupet & Chantraine, 1992; Hupet et al., 1993, 1991; Isaacs & Clark, 1987; Wilkes-Gibbs & Clark, 1992).

A different pattern of results was obtained when both content words and non-content words were included in the analysis, and also in the analysis on the number of hedges produced. The total number of words produced by Directors was reduced in critical trials, but this was true regardless of whether or not the pictures discussed were in the partners' common ground. Likewise, the number of hedges produced by Directors was reduced in critical trials regardless of condition the pair was in. The effect of Target Type on the total number of words produced could reflect description production being facilitated by the Director's prior knowledge of the referent regardless of common ground. As for hedges, it is noteworthy that previous studies have found that the reduction in the number of hedges produced after a referential expression has been added to the common ground is stronger when participants are given more opportunities to refer to a picture. For instance, in Brennan and Clark's (1996) study, participants had the opportunity to discuss pictures either once or

four times, thus adding the corresponding referential expressions to their common ground. The authors found that when these participants referred to these pictures again, they produced fewer hedges in their descriptions, but that this was mainly the case when they had had the opportunity to discuss the pictures four times previously. Recall that in the current study, the participants only had the opportunity to discuss the figures once during the Dialogue Phase; thus, one possibility is that the corresponding memory representation was not strong enough to affect hedge production during the Matching Phase (it was, however, strong enough to affect the production of content words). In any event, the fact that the number of hedges produced also depended on Target Type regardless of Condition suggests that hedge production depends at least in part on the speaker's knowledge, regardless of whether the referent under discussion belongs to the common ground or not.

Taken together, these results are consistent with the idea that prior knowledge of the pictures shown during the Matching Phase affected the Director's speech. Whilst some aspects of the Director's speech were affected regardless of whether or not these pictures belong to the common ground (i.e., total number of words produced, number of hedges produced), other aspects were also sensitive to whether or not the Tangram figure under discussion was also known to the Matcher (i.e., number of content words produced), confirming that the common ground built during the Dialogue Phase affected at least part of the descriptions produced during the Matching Phase.

Appendix B: Main analysis – Full random effects model output

Table B1

Covariance Parameter Estimates

Parameter	Level	Estimate	SD
Random intercepts	Dyad	0.00	
Random slopes – number of referential expressions initially considered	Dyad	0.00	
Random slopes – generation	Dyad	0.00	
Random slopes – frequency in the corpus	Dyad	0.00	
Random intercepts	Participant	0.37	0.14
Random slopes – number of referential expressions initially considered	Participant	0.00	
Random slopes – generation	Participant	0.00	
Random slopes – frequency in the corpus	Participant	0.00	
Random intercepts	Item	0.15	0.16
Random slopes – condition	Item	0.08	0.14
Random slopes – number of referential expressions initially considered	Item	0.05	0.16
Random slopes – generation	Item	0.00	
Random slopes – frequency in the corpus	Item	0.15	0.12

Table B2

F Values

Effect	DFs	<i>F</i>	<i>p</i>
Condition	1, 60	1.21	.276
Number of referential expressions considered	1, 56	8.76	.005

Generation	1, 909	15.78	< .001
Frequency in the corpus	1, 60	50.55	< .001
Frequency in the corpus x Condition	1, 495	5.20	.023

Table B3

Model Parameters

Effect	Estimate	Standard error	<i>p</i>
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		

Appendix C: Analysis of the reuse of non-preferred referential expressions

The hypothesis tested in this study focused solely on the reuse of the participants' preferred referential expressions. However, the participants' dispreferred referential expressions were also coded, and their reuse was analysed for information purposes only. This analysis is reported hereafter and was conducted following the same rationale as the main analysis. The only difference is that the IV "Number of referential expressions initially considered" was not included in the analysis, because all dispreferred referential expressions were necessarily considered in trials where more than one referential expression was initially considered during the Dialogue Phase.

The mixed model used to analyse these data included Condition, the identity of the participant who initially generated the referential expression and the referential expression's frequency in the corpus as fixed effects. The outcome variable was the probability of the Director reusing one of the dispreferred referential expressions during the Matching Phase. The random effects structure included by-dyad and by-participant random intercepts, by-dyad random slopes corresponding to the referential expressions' frequency in the corpus and by-item random slopes corresponding to Condition. The data are shown in Table C1.

Table C1

Proportion of Matching Phase Trials where the Dispreferred Referential expression was Reused as a Function of the Condition and of Who Initially Generated the Referential Expression

	Same Partner	Different partner	Grand total
Self-generated	0.23	0.27	0.25
Partner-generated	0.16	0.20	0.18
Total	0.19	0.24	0.21

Firstly, a significant generation effect was found, $F(1, 551) = 4.46, p = .035$. Directors were more likely to reuse a dispreferred referential expression when they had initially generated it themselves than when their partner had initially generated it, $OR = 1.62, CI_{.95} = 1.04, 2.55$. Secondly, a significant effect of frequency in the corpus was found, $F(1, 40) = 28.40, p < .001$. The likelihood of Directors reusing a dispreferred referential expression increased as the referential expression's frequency in the corpus also increased, $b = 0.97$. Finally, the main effect of Condition failed to reach statistical significance, $F(1, 38) = 0.40, p = .530$. In sum, the reuse of dispreferred referential expressions depended mainly on two factors: the identity of the participant who had initially generated them and their perceived relevance in the corpus.