



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# The emergence and adaptation of referential conventions in dialogue

Lucia Castillo



Degree of Doctor of Philosophy

Psychology

School of Philosophy, Psychology and Language

Sciences

University of Edinburgh

2019



## Declaration

I hereby declare that this thesis is my own composition and that the work reported here has been carried out by myself (except where due acknowledgement is made in the text). This work has not been submitted for any other degree or professional qualification.

The following chapter of this thesis is a formatted manuscript that has been submitted to peer-reviewed journals:

Chapter 3: Castillo, L., Smith, K. & Branigan, H. Submitted to *Cognitive Science*, currently under revision to be re-submitted.

Authorship details: Lucia Castillo designed and ran the study, analysed the data, and wrote the original manuscript. Holly Branigan and Kenny Smith acted as supervisors, gave feedback on each step of the process, and contributed to the revision of the manuscript.

A handwritten signature in black ink, appearing to read 'Lucia Castillo', with a long horizontal stroke extending from the top of the signature.

Lucia Castillo

7th January 2019



## Abstract

When speakers in dialogue are faced with the need to repeatedly refer to the same items, they usually use the same references they or their partners had used before. These previously-used references act as precedents, standing in speakers' memory as successful ways of solving that particular communicative need. From mechanistic models explaining this reuse as a consequence of low-level priming (Pickering & Garrod, 2004), to models assuming sophisticated partner-modelling processes (Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986), most theories assume speakers will maintain their referential choices throughout the dialogue. While references might be modified to discard superfluous elements (as in *referential reduction*, Krauss & Weinheimer, 1964; or *simplification*, Bard & Aylett, 2005), conceptualisations will be preserved. However, there are several reasons why speakers might need to change their previous referential choices: a context change might render the old reference insufficient to identify the target, or overly detailed, prompting the listener to wonder if additional meanings are implied. It might also be that the repetition of the task highlights a better referential alternative, or that additional information makes other alternatives more salient for the speakers. In this thesis, I present an investigation of the dynamics of reference repetition and change in dialogue, bringing together a theoretical analysis of the existing literature and 5 experiments that aim to clarify how, when, and why speakers might change their referential choices.

Experiment 1 explores the dynamics of reference change when the repetition of a task creates additional pressures that were not evident in an initial exposure. The experiment compares pairs and individual speakers describing positions in two spatial contexts (regular or irregular mazes, as in Garrod & Anderson, 1987; Garrod & Doherty, 1994) that cue participants into using different referential descriptions. As the task is repeated over 3 rounds, an additional pressure for efficiency is created, pushing participants across contexts into using one of the two initial descriptive choices. Crucially, only interacting pairs of participants adapted to this additional pressure by switching to a more efficient alternative, while participants completing the task individually maintained their initial choices. Additionally, this chapter reports a pilot study of the same experiment in 4-person groups that showed a switching pattern similar to the one found in interacting pairs.

Experiments 2 and 3 further investigate the drivers of linguistic change in referential choice and the relationship between interaction and adaptation found on experiment 1. Experiment 2 explores the relationship between context change and linguistic change. Using the maze game paradigm, the experiment presented individuals and pairs of participants with either the same maze in each round of the task, or different mazes in the first and the second halves of the experiment. The results of Experiment 2 offer some support to the conclusions of Experiment 1, as participants switched to Abstract descriptions as they gained experience in the task; however there were no significant differences between Interaction conditions, nor between Same or Different mazes. Experiment 3 was aimed at exploring which features of interaction were relevant for reference change. The experiment used the maze game in different interactive setups, in which participants played a first round of the game either as Matchers (in direct interaction with the Director) or Overhearers (having access to another pair's dialogue), and three successive rounds as Directors with either the same partner as in the first round, or a new partner. Participants showed higher levels of adaptation as they gained experience in the task; while the different interactive setups did not significantly influence their reference choices.

Experiments 4 and 5 further explore the relationship between reference change and participant role in interaction. Using a picture matching paradigm (Brennan & Clark, 1996), Experiment 4 tested participants interacting with either the Same Partner throughout, a New Partner in the second half of the experiment, or an Overhearer (who had witnessed the first half of the experiment) in the second half. Participants in all conditions maintained previously used overspecific picture descriptions even when those detailed descriptions were not needed to identify the referents, pointing towards a predominance of speaker-centred factors in reference choice. Experiment 5 used a similar interactive setup, testing participants on a larger set of pictures. Participants maintained their overspecific descriptions only if interacting with the Same Partner they had on the first half of the experiment, or with the Overhearer, switching to context-appropriate basic-level descriptions if interacting with a New Partner. Taken together, both experiments suggest a complex balance between speaker-centred and audience-design factors in the potential change of reference choices, where speakers need to weigh their own effort against the communicative needs of their partner.

These experiments highlight the crucial role of interaction in the adaptation of reference choices to changes in context, and show that individuals' 'conservative bias'

that leads them to maintain their own previously used references can be overturned in the search for better communicative alternatives in interactive dialogue.



## Lay summary

People tend to reuse the same labels that have already been used in a conversation: their likelihood of using “*the Psychology building*” or “*7 George Square*” depends on which option their partner (or themselves) has used before. As this label is active in the speaker’s memory, it is not surprising it comes as an easier option when facing the need to refer to the same object or event. Indeed, most theories of dialogue suggest this repetition is crucial for understanding between interlocutors. However, there are circumstances in which the speaker will need to abandon this previously used label for a different option. A change in context, or a need to differentiate this referent from similar items, might push the speaker to adapt to these new communicative needs, switching from a known label to a new one. This thesis looks at both sides of this issue: the reuse of labels in conversation, and the adaptation process that allows speakers to change when a different alternative is needed.

The first aim of this thesis was understanding the role of the context in the reuse or adaptation of referential choices. To look at this issue, I manipulated the perceptual context that speakers had to discuss, either making some aspects of the context more salient than others (Experiment 1), or having speakers refer to slightly different contexts in different rounds of a task (Experiment 2). Moreover, these experiments compared individual speakers, producing descriptions for an imagined future partner, and pairs of speakers in interactive dialogue, in order to understand how interaction affects reference use. The experiments’ results suggest speakers initial reference use is influenced by salient aspects of the context, but that they will abandon this initial choice to adapt to the communicative task, a process that is facilitated for speakers in interaction. A second set of experiments (Experiments 4 and 5) asked how a change from a more complex to a simpler context affected the likelihood of participants maintaining their original references (that were too specific in the new context), or switching to more basic labels, thereby adapting to the context change. These experiments showed speakers would preferentially retain the old labels, but may change them if this appears like an efficient choice.

The second aim of this thesis was understanding how the different roles an individual can play in interaction affect the referential adaptation process. We usually think about dialogue as an interaction between two people, both of whom have full access to what is being said; but dialogues can have other participants, either involved in the

interaction but not directly addressed (side-participants), or not recognised as participants at all (overhearers). We asked how either being in one of these peripheral roles (Experiment 3), or talking to someone that was previously an overhearer (Experiments 4 and 5) affected the likelihood of retaining previously used references. The results suggest that speakers consider these peripheral roles similarly to how they consider their actual addressee, opting to retain previously used labels unless they were interacting with a completely new partner (and even then, they would adapt only in specific circumstances).

Taken together, the experiments presented in this thesis suggest speakers are conservative when it comes to labels, but that they are always evaluating the fit between their reference choices and the actual circumstances of use of these references. This would explain why interaction facilitates the adaptation process, as the feedback they get from their partner can help them confirm the change was a good decision.



## **Acknowledgements**

This thesis has been completed with funding from CONICYT, the Chilean agency for Science and Technology, under a Becas Chile scholarship (2013 – 72140201).

First of all, I would like to thank my supervisors, Holly Branigan and Kenny Smith, for their unconditional support and patience during these four years. Not only they have been wonderful in terms of their academic support, but their personal qualities and the genuine care they have always expressed have made the most stressful period of my life so much bearable. Additionally, I would like to acknowledge Martin Pickering; his comments and the interest he has always shown in my work have had a positive impact on this thesis. The insightful comments of my two examiners, Chris Cummins and Simon Garrod, also contributed significantly to this work, and made for an interesting and enjoyable (!) viva.

The work of these four years has been greatly improved by the countless lab meetings of the Psycholinguistics group, whose past and present members have all contributed to my academic development. The now defunct Joint Action Reading Group (JARG) was also a wonderful space to discuss and think outside the box. Special thanks go to Alessia Tosi, Chiara Gambi, and Jarek Lelonkiewicz, who kept it going for as long as they possibly could.

On a more practical level, I would like to thank Alisdair Tullo for making the experiments work, and for making me believe I could actually write some code. The DiET chattool, created and maintained by Greg Mills, served as the basis for the software written for this thesis.

A few special people have managed to be friends, colleagues, and counsellors, all at once: my deepest gratitude goes to Carmen Saldaña, Marieke Woensdregt, and Anita Tobar; and to my deeply missed Pablo San Martín, now back in Chile. I'm not exaggerating when I say I could not have done this without you.

To my husband Ricardo and my daughter Noelia, to my family and friends: Your love has kept me going. I promise this is the last thesis I'll write!



## Contents

<b>1. Chapter 1: Introduction .....</b>	<b>1</b>
1.1 Repetition and change in reference production.....	3
1.2 Individual and interactive processes in dialogue .....	4
1.3 Thesis aims .....	6
1.4 Overview .....	7
<b>2. Chapter 2: Literature review.....</b>	<b>10</b>
2.1 Imitation and convergence in communication .....	11
2.2 The production of references in dialogue: Historical and ahistorical factors ...	14
2.3 Common ground.....	22
2.4 Interaction and participant role in conversation.....	27
2.5 Conventions and change.....	30
2.6 Relationship between selection at the dialogue level and language change ..	36
2.7 Conventionalisation in non-linguistic communication .....	40
<b>3. Chapter 3: Interaction promotes the adaptation of referential conventions to the communicative context.....</b>	<b>44</b>
3.1 Introduction .....	44
3.2 Method .....	54
3.3 Results .....	60
3.4 Discussion.....	71
3.5 Conclusions.....	80
3.6 Appendix: Emergence and adaptation of referential conventions in small communities: Introduction.....	83
3.7 Method .....	84
3.8 Results .....	86
3.9 Discussion.....	93
<b>4. Chapter 4: Drivers of reference change .....</b>	<b>95</b>
4.1 Introduction .....	95

4.2 Experiment 2: Relationship between context change and linguistic change...	97
4.3 Experiment 3: Participant role in referential adaptation.....	114
4.4 Conclusions.....	127
<b>5. Chapter 5: Weighing of production factors in the use of referential precedents</b> .....	<b>129</b>
5.1 Introduction .....	129
5.2 Experiment 4 .....	146
5.3 Experiment 5 .....	161
5.4 General discussion.....	172
5.5 Conclusions.....	176
<b>6. Chapter 6: Conclusions</b> .....	<b>178</b>
6.1 Summary.....	180
6.2 Discussion.....	182
6.3 Limitations and implications for future research .....	183
<b>References</b> .....	<b>186</b>

## List of figures and tables

### Figures

3.1	Experiment 1: Mazes for player A and player B .....	51
3.2	Experiment 1: Regular and irregular mazes .....	55
3.3	Experiment 1: Tangram set .....	56
3.4	Experiment 1: Proportion of descriptions using Concrete or Abstract schemes per round .....	62
3.5	Experiment 1: Time taken to complete experiment, Pairs and Individuals .....	65
3.6	Experiment 1: Proportions of non-aligned and aligned descriptions .....	67
3.7	Experiment 1: Mean number of words per tangram per round .....	70
3.8	Appendix: Mean description scheme use per Group .....	87
3.9	Appendix: Mean alignment per Group .....	89
3.10	Appendix: Mean tangram description length per Group .....	91
3.11	Appendix: Mean time taken to complete each round, per Group .....	92
4.1	Experiment 2: Irregular and regular mazes .....	100
4.2	Experiment 2: Tangram set .....	100
4.3	Experiment 2: Proportion of descriptions using Concrete or Abstract schemes per block .....	104
4.4	Experiment 2: Mean use of Abstract or Concrete description schemes per round .....	104
4.5	Experiment 3: Irregular and regular mazes .....	117
4.6	Experiment 3: Regular maze from Director and Matcher's point of view ...	118
4.7	Experiment 3: Distribution of participants per condition .....	120
4.8	Experiment 3: Proportion of description use per round .....	123
4.9	Experiment 3: Proportion of aligned descriptions .....	125
5.1	Experiment 4: Set A and Set B examples .....	147
5.2	Experiment 4: Chat displays for Director and Matcher .....	148
5.3	Experiment 4: Task setup .....	150
5.4	Experiment 4: Mean use of the same description, rounds 4 to 6 .....	154
5.5	Experiment 4: Mean use of basic-level labels, rounds 4 to 6 .....	155



5.6	Experiment 4: Mean use of reduced descriptions, rounds 4 to 6 .....	156
5.7	Experiment 4: Proportion of overspecification type, round 4 .....	157
5.8	Experiment 4: Frequency of marker use, round 4 .....	158
5.9	Experiment 5: Set A and Set B examples .....	162
5.10	Experiment 5: Example of image set A, from Matcher's view .....	163
5.11	Experiment 5: Mean use of the same description, rounds 4 to 6 .....	165
5.12	Experiment 5: Mean use of basic-level labels, rounds 4 to 6 .....	166
5.13	Experiment 5: Mean use of reduced descriptions, rounds 4 to 6 .....	167
5.14	Experiment 5: Frequency of marker use, round 4 .....	168

## Tables

3.1	Experiment 1: Results of model of description scheme use.....	63
3.2	Experiment 1: Results of model of time taken to complete task, Pairs condition.....	64
3.3	Experiment 1: Results of model of time taken to complete task, Individual condition .....	65
3.4	Experiment 1: Results of model of aligned descriptions .....	68
3.5	Experiment 1: Results of model of tangram description length .....	71
3.6	Appendix: Organisation of interacting pairs per round .....	86
3.7	Appendix: Results of model of description scheme use .....	88
3.8	Appendix: Results of model of aligned descriptions .....	90
4.1	Experiment 2: Mean use of Abstract descriptions and other descriptive statistics.....	103
4.2	Experiment 2: Results of model of description scheme use.....	106
4.3	Experiment 2: Mean use of Abstract descriptions, first round .....	107
4.4	Experiment 2: Results of model of aligned descriptions .....	109
4.5	Experiment 2: Mean use of aligned descriptions .....	110
4.6	Experiment 3: Mean use of Abstract descriptions and other descriptive statistics .....	121
4.7	Experiment 3: Results of model of description scheme use .....	122
4.7	Experiment 3: Results of model of aligned descriptions .....	124



## Chapter 1. Introduction

The last few decades of Psycholinguistics research have established the fact that people repeat each other in dialogue well beyond doubt. We tend to use the same labels our interlocutor has used to refer to the same objects or events, and we reuse the same syntactic structures, even more if similar semantic content is involved (Branigan, Pickering, McLean, & Cleland, 2007). Interacting speakers even tend to pronounce words similarly (Pardo, 2006). The whole business of using language in interaction seems to be defined by convergence and repetition, and, indeed, the evidence suggests repetition eases the effort of communication (Pickering & Garrod, 2004).

But this repetition has obvious limits. How else would we suggest a different conceptualisation, or mark a new perspective, or adapt to a new communicative situation, if not by introducing a *change*? The goal of this thesis is to study the relationship between repetition and change in the production of references in interaction. Specifically, my research aims at understanding referential adaptation, defined as the process by which speakers abandon a repeatedly-used reference to switch to a new formulation that is deemed a better fit for the communicative situation, usually entailing a different conceptualisation. In the chapters that follow, I will address some of the main components of this process: the role of context in the emergence and adaptation of referential conventions, the relationship between interaction and adaptation, and the modulation of reference use by the participant role of the speakers. While accepting that repetition accounts for most of the regular use of references in communication, I believe explaining how speakers manage to deviate from this norm can give us relevant insights on the nature of reference production and use.

The idea that reference change in linguistic interaction might be seen as a window into the general process of language change is central to this thesis. In particular, as the selection of references can be understood as an attempt to find the alternative that best fits the communicative situation, the study of referential change in interaction should be able to inform more global processes of selection of linguistic and cultural variants in a population. Even though most of the processes studied under the 'language change' label can be described as unconscious (the accumulation of phonetic changes in the articulation of frequent words, the repetition of syntactic structures, etc.), the conscious choices of the speaker when producing a reference might also be relevant when studying the differential accumulation of lexical variants in a population over time.

This thesis is also defined by a joint action view of dialogue, which understands the individual contributions of each speaker in a conversation in the context of the joint communicative act that is taking place between them. Moreover, I assume that this joint action is built from both intentional and unintentional aspects of communication; that is, that while the overall facts of the interaction (such as topic, participants, language, communicative intentions) can be said to be intentional, many other aspects are likely unintentional (such as tone, gestures, syntactic choices), with both shaping the actual linguistic output of the speakers (Garrod & Pickering, 2009; Tollefsen & Dale, 2012). Similarly, even though the mechanisms of joint action might involve both 'low-level' perception-action processes, and 'high-level' cognitive processes, it is likely that both levels work in integration (Galantucci & Sebanz, 2009).

In the same sense, the appeal to the consideration of the perceptual context in the production and comprehension of references is, even though 'obvious' in the sense that references point to a meaning in the world, relevant theoretically as context is intrinsically linked to the joint action being performed. As several of the experiments presented in this thesis will show, the evaluation that speakers perform of their context in dialogue is incorporated to their language production and comprehension processes, both intentionally and unintentionally. Even though there is a fundamental difference in the interactive coupling of individuals in joint action and their individual connection to the world, both aspects are inescapable in the analysis of dialogue as a situated process.

## 1.1 Repetition and change in reference production

Besides the most abstract elaborations, most instances of linguistic communication involve a speaker trying to get an addressee to attend to a specific element in their common context. In order to specify which, of all possible items and conceptualisations that even the most sparing context has to offer, is the one she intends to guide the addressee's attention to, the speaker needs to come up with a referential expression. Such a reference should ideally offer enough information for the addressee to identify the intended meaning without having to incur an extra turn to confirm it, and should not include too much information so as to make the addressee suspicious that other meaning is actually being considered. Now, their common context is not only formed by their material surroundings; it includes also their social and cultural memberships, their common history, and, crucially, *their previous linguistic interaction(s)*, up to the very moment of production of the next utterance.

A reference that has been proposed and accepted in dialogue will likely be repeated by both interlocutors, who will preserve its conceptual content while eliminating superfluous elements and generally streamlining its surface form. While different theories of dialogue have explained this repetition by appealing to more mechanistic processes like priming and entrainment, or more conscious processes like audience-design, it is commonly acknowledged that reusing their partner's lexical and syntactical choices is both a 'safe bet' in terms of communication (as they are obviously understood by the first speaker), and an easy choice in terms of processing (as they were already active in both speakers' memory). In reusing a precedent, speakers communicate their accepting of this choice, and through this mutual ratification, a local convention is established.

However, as we suggested previously, repetition cannot account for the whole of reference use, as there are many instances where the communicative situation would cue the need for a change. A change in context might render a previously-used reference underspecific, prompting the speaker to expand it or change its specification level, e.g., going from "*the dog*" to "*the white terrier*" when other dogs are included in the context. Repetition of a task might push speakers to find a more efficient alternative, e.g., moving from idiosyncratic descriptions like "*two along and*

*then two up, at the very corner*” to a system of coordinates like “*row 2, column 3*”. An overly detailed description might be abandoned if the context is simplified, either reducing accessory elements or switching to a different specification level. A new task-prompted perspective on an object might push speakers into abandoning a previously used label associated to a different conceptualisation (Ibarra & Tanenhaus, 2016).

Considering this, it might be thought that referential adaptation is a process determined solely by the speaker, who chooses to either maintain their previous reference or switch to a new one. However, on the one hand there are many aspects of the production process that occur below conscious control. Even those aspects that might be considered explicit might escape speakers’ awareness until the very moment of utterance (including the ‘decision’ to change in the first place). In this sense, even though the process is carried out by the speaker, it cannot be said to be completely determined by her.

On the other hand, speakers in dialogue mutually influence each other’s choices not only through their linguistic output, but through a myriad of intentional and unintentional behaviours that are interpreted in an integrative fashion. In this process, the frontiers of what comes from the inside, as egocentric processing needs or cognitive constraints, and from the outside, as explicit feedback, hedges, subtle changes and even the time-course of a response, might be blurred. Even in dialogues with no face-to-face interaction, such as the ones described in this thesis, speakers manage to influence each other in a range of levels, and even silences (or delays, in written chat communication) can be said to be included in their dynamic evaluation of the communicative situation.

## **1.2 Individual and interactive processes in dialogue**

As Schober and Clark (1989) stated, the social process of interacting in conversation plays a central role in the cognitive process of understanding. But how much of the outcome of a communication event can be explained by interactive vs individual processes is a hotly debated matter. Some accounts have favoured egocentric processing as the main force behind both the production and

comprehension of references, showing how individuals will prioritise their own perspective when interpreting or designing an utterance, even when the perspective of their interlocutor would point to a different meaning (Keysar, Lin, & Barr, 2003; Lin, Keysar, & Epley, 2010). These accounts argue that audience-design, or perspective taking, is a costly process that is not really needed in most communicative situations, as speakers are usually able to take their own perspective as proxy for their interlocutor's, considering common factors such as a common context, cultural group membership, etc.

Other accounts have focused on the integration of speaker-centred and partner-specific processes, claiming that the addressee's knowledge and needs (or the speaker's, in comprehension) can be taken into account from the earliest moments of processing. These accounts emphasize the integration of both sources of information, however, the discussion they present is also centred on the time-course of processing, reflecting a compromise with a linear view of information integration (Brennan & Hanna, 2009; Hanna, Tanenhaus, & Trueswell, 2003).

A recent view proposing that the integration of sources of information can be best explained through a dynamical systems approach might be closer to the epistemological stance of this thesis. Dale et al. (2018) propose a model where the information from sources acting on different timescales (such as 'fast' egocentric processing and 'slow' other-centred processing) is integrated in a flexible and task-relevant way. As the authors explain, an other-centred perspective can come after an egocentric perspective in time, but its pervasiveness might be crucial in determining the communicative outcome. In this sense, the traditional discussion that tries to determine a hierarchy of processes might be misled, focusing on 'what comes first' instead of on the most stable patterns of outcome. Degen and Tanenhaus (2016) similarly call for a dynamical integration of sources that argues that the idea of 'fast' and 'slow' processes is not really relevant when we consider production and comprehension are situated in a context that might differentially guide speakers' evaluation of the communicative situation.

In this thesis, I look at the linguistic choices of speakers in interaction (or in individual settings with a communicative intent), with the aim of establishing the relative importance of factors affecting their production: the perceptual context, the

participant role their partner has previously had, the interactiveness of the task. However, I believe that none of these factors can be completely isolated analytically, as a robust interpretation of their contribution is only possible if we acknowledge all other elements the speakers are having to consider both intentionally and unintentionally.

In the experiments summarised in the following chapters, I address the factors involved in the production and adaptation of references using two classic paradigms in psycholinguistics: the maze game (Garrod & Doherty, 1994; Garrod & Anderson, 1987), and the picture-matching task of Brennan and Clark (1996). Both paradigms have been described as prototypical tasks in the language-as-action tradition (Trueswell & Tanenhaus, 2005); however, their instantiation in written form through the use of chat-based software makes additional layers of information available to the researcher that are not so easily extracted from regular conversation, such as detailed time-stamps of production and turn-taking, pairing of task movements (mouse clicks and maze navigation) and linguistic production, etc. Even though the scope of this work does not allow for a detailed analysis of all these features, the integration of new digital tools and classic communication paradigms opens new possibilities for analysis, carefully controlling interaction channels while maintaining the rich environments and situations that this tradition has supported. While it might be criticised that written chat communication does not allow for the full spectrum of communicative signals that are exchanged between individuals in face-to-face interaction, it allows us to focus exclusively on their linguistic output, eliminating the need to control for other sources of information (such as looks, or subtle tone changes) that might partially explain speakers' behaviour in other situations.

### **1.3 Thesis aims**

This thesis aims to shed light over the process of referential adaptation, by which speakers abandon previously used references and switch to a different alternative in order to adapt to some feature(s) of the interaction context. Specifically, I address the influence of three main factors over this process: the perceptual context (the scenario in which the speakers' references are to be interpreted), the interaction situation (whether the speakers are interacting directly with a partner), and the



participant role of the interlocutors (whether they are interacting continuously with the same partner, or with a new partner, or with other individuals with different degrees of involvement in the interaction). The general goal of the research project is to try to understand what leads people to abandon entrained lexical choices, in order to shed light on the process of conventionalisation (or routinisation) of linguistic alternatives in a community, and how this relates to the overall process of language change.

#### **1.4 Overview**

This thesis describes 5 experiments that deal with the issues surrounding the emergence and adaptation of referential choices, together with an analytical summary of the relevant literature. Chapter 2 presents a literature review, aimed at clarifying what are the factors that speakers consider when proposing a reference, and what are the conditions under which they would be inclined to change it. It also covers the impact different production circumstances might have on this process, particularly the participant role of the speaker; and establishes a link between change at the psycholinguistic level –the focus of this thesis– and language change.

Chapter 3 reports the first experiment, aimed at understanding the role that the perceptual context plays in the choice of referential descriptions, and at establishing the differences in processing of speakers working individually vs in interactive dialogue. Using a maze game paradigm, the experiment measured pairs and individual speakers in their use of abstract and systematic or concrete and figurative descriptions to identify a position in the maze, along with their alignment in description scheme use and speed in the task. Participants' choice of descriptions was found to be affected by the perceptual context, while interaction appeared as crucial to overcome contextual effects in order to switch to a more efficient alternative. An appendix to this chapter includes the results of a pilot experiment that applied the same design to small communities of speakers.

Chapter 4 presents Experiments 2 and 3, which used the maze game paradigm aiming to understand which aspects of interaction are most relevant in referential adaptation and change. Experiment 2 explored the influence of a context change

over the adaptation of referential choices, presenting individuals and pairs of participants with either the same maze layout in each round of the task, or different mazes in the first and second halves of the experiment. Results show this manipulation was not enough to drive different patterns of responses in the experimental conditions, with participants switching to more adaptive descriptions as they played more rounds in the task, across conditions. Chapter 3 explored the effect of participant role on referential adaptation, using the maze game paradigm while manipulating the role participants had to perform (Matcher or Overhearer) and the identity of their partner (Same Partner vs New Partner). The experiment showed that again participants across conditions switched to the more efficient Abstract descriptions as they gained experience in the game, pointing to the need for stronger experimental manipulations that could overcome the generalised influence of experience on referential adaptation.

Chapter 5 describes Experiments 4 and 5, which used a picture-matching game (Brennan & Clark, 1996) to address the interplay between speaker-centred and audience-centred factors in the production of references. In the task, the pictures described in the first half of the experiment are shown in a different context in the second half, which renders the references previously used overspecific in this new context. Experiment 4 had participants playing with either the same partner throughout the task, a new partner in the second half of the task (who had not witnessed their interaction in the first half), or an overhearer (who had witnessed their interaction in the first half). Participants across conditions maintained their overspecified descriptions in the second half of the game, pointing towards a predominance of speaker-centred factors in reference choice. Experiment 5 used a larger, monochrome set of images with the same participant conditions, and found that participants maintained their previously used descriptions when interacting with the same partner or the overhearer who had witnessed their previous interaction, but switched to more appropriate basic-level references when interacting with a new partner. Taken together, the results of both experiments point to a dynamic evaluation of the factors influencing reference production, where speakers take into account both their own processing needs (crucially modulated by task context) and their partners'.

The closing Chapter 6 presents the overall conclusions from this body of work, considering the evidence presented by the experiments, and critically assessing the points that could be improved in future research, both in terms of methodological challenges and hypothesis specification.

## **Chapter 2. Literature review**

When a speaker chooses a reference, several factors come together: the influence of frequency and recent use over the availability of lexical alternatives, the perceptual and linguistic context, the identity of her interlocutor and their common history of interaction, her communicative goals, etc. The goal of this chapter is to describe how each of these elements affects the production of references in interaction. Specifically, the chapter will cover the factors that influence initial reference choice, the dynamics of conventionalisation in repeated use, and the factors that usually motivate reference adaptation and change. Additionally, it will briefly review the state of the art on the link between selection at the dialogue level, and language change.

Section 2.1 will cover the role of imitation and convergence in communication, describing briefly the two main models explaining repetition in dialogue: the interactive alignment model (Pickering & Garrod, 2004), and the collaborative model of H. H. Clark and colleagues. Section 2.2 will discuss the production of references in dialogue, specifically the use of the linguistic and non-linguistic context in this process. Section 2.3 will focus on common ground and coordination processes in dialogue. Section 2.4 will discuss the differences in participant role in conversation. Section 2.5 will address the establishment of linguistic conventions and their consolidation or change; and section 2.6 will deal with the link between convention formation in dialogue and language change. Additionally, section 2.7 will address the similarities and differences between adaptation and change in linguistic vs non-linguistic communication, looking specifically at experimental semiotics studies.

## 2.1 Imitation and convergence in communication

Imitation is widespread in human behaviour. Human beings tend to imitate each other during social interaction, at many levels. This behaviour ranges from the imitation of facial expressions and body postures (Chartrand & Bargh, 1999; Louwerse, Dale, Bard, & Jeuniaux, 2012; Shockley, Santana, & Fowler, 2003), to syntactic and lexical repetition in conversational partners (Branigan, Pickering, & Cleland, 2000; Pickering & Garrod, 2004, 2013). While the specific mechanisms that explain each of these behaviours might involve both conscious and automatic processing, it has been suggested that most imitative behaviours are not consciously controlled by those who perform them (Chartrand & van Baaren, 2009; Heyes, 2011).

Imitation has been linked to benefits in different areas of social interaction. On an affective level, imitation has been associated to increased affiliation (Chartrand & Bargh, 1999; Lakin & Chartrand, 2003; van Baaren, Holland, Steenaert, & van Knippenberg, 2003); a finding that has been replicated in non-human primates (Paukner, Suomi, Visalberghi, & Ferrari, 2009). Imitation can also aid in linguistic comprehension. Adank, Hagoort, & Bekkering (2010) showed that imitating the accent of an interlocutor aids in the comprehension of spoken sentences. Imitation can make it easier to predict subsequent actions of the imitated subject, both in terms of motor sequences (Wilson & Knoblich, 2005) and language (Pickering & Garrod, 2007), and overall, it reduces the effort of dialogue, as the items that are processed in the comprehension of an utterance are the same items that will be employed in production, demanding no additional activation (Pickering & Garrod, 2004).

Imitation in natural dialogue has been extensively studied, particularly in order to explain speakers' tendency to repeat each other's choices in terms of words, syntactic structures, conceptual schemes, etc. Two main theoretical standpoints can be identified: 1) **a mechanistic approach** around the *alignment model* (Pickering & Garrod, 2004), proposing that this convergence is reached through mechanistic priming on the most basic levels (speech rate, word similarity, syntactic structures, lexicon, etc.) that "percolates" towards more complex levels (semantic, situation model) without guided intervention; and 2) **a collaborative approach** focused on

the interactive coordination of speakers and usually requiring some form of intentionality<sup>1</sup>. In this category, I draw together both the classic common ground approach of Clark and colleagues (Clark & Brennan, 1991; Clark, 1996) with the more recent *interpersonal synergy* models of Fusaroli and colleagues (Fusaroli, Raczaszek-Leonardi, & Tylén, 2014; Fusaroli & Tylén, 2012).

The alignment model (Pickering & Garrod, 2004, 2006, 2013) states that a shared interpretation of the set of concepts and ideas that are being discussed (that is, a common understanding of the referents that are being discussed, even if *perspectives* on those referents are not shared), that is, an *alignment of situation models*, forms the basis of successful dialogue. This high-level alignment is built up from mechanistic alignment in lower levels, such as syntax, tone, speech rate, etc., where the imitation of the interlocutor's output *via* priming plays a key role. As such, the whole process rests on a principle of input-output coordination, where the linguistic output of the interlocutor primes the speaker, leading to the reuse of the same items in their own turn. This reuse mechanism eases the effort of the conversation, since the items that were used in the comprehension of the previous utterance are already active in processing when they are reused in production, avoiding the need for new activations. In this sense, the basic mechanism of alignment is automatic (Garrod & Pickering, 2007), even though it can be modulated by speakers' beliefs (Branigan, Pickering, Pearson, McLean, & Brown, 2011). The authors argue that this low-cost mechanism is responsible for most of successful dialogue, with more sophisticated and costly mechanisms (such as audience design) only coming in place if understanding cannot be reached through this simpler process.

The collaborative approach can be defined around the idea of dialogue as a joint action that is consciously approached by the interlocutors, where imitation aids in the effort of reaching a joint understanding of the concepts and ideas being discussed. In these models, imitation occurs as a consequence of the need for mutual understanding, in the sense that the repetition of a linguistic item both eases the effort of the conversation (according to the principle of *least collaborative effort*,

---

<sup>1</sup> The Speech Accommodation theory (Giles, Coupland, & Coupland, 1991) shares with the collaborative approach an intentional stance, where speakers are actively modulating their output to accommodate to the perceived needs of their interlocutors. However, as it is focused on social aspects of communication that are beyond the scope of this thesis, it won't be discussed further.

Clark & Wilkes-Gibbs, 1986) and ensures to the conversational partner that the meaning of the item is shared, and will from then on be considered as part of their *common ground*. As the idea of common ground requires the information that is shared to be mutually known to be shared, using the same linguistic items the interlocutor just used communicates both the acceptance of that reference and its adoption as shared between the speakers. Contrary to the alignment model, the collaborative model requires that speakers maintain a model of their addressee's state of mind, with both interlocutors working together to ease the effort of the conversation.

The idea of *interpersonal synergy*, central to the most recent developments of this approach, puts the emphasis not on imitation but on complementary performances by interdependent interlocutors, which may employ imitation but are not defined by it (Fusaroli & Tylén, 2016). While this approach considers coupling between the interlocutors (in terms of their interdependent linguistic and non-linguistic joint actions), *interpersonal synergy* should not be considered an automatic phenomenon, as the complementarity of the speakers is dependent on the communicative intention and semantic/pragmatic content of their utterances.

Though much of the experimental support for imitation models in language comes from 'one-off' tasks, where participants are primed to reuse a lexical item or syntactic structure by either another participant or a confederate having used the same structure previously (see for example Branigan et al., 2000; Branigan, Pickering, McLean, & Cleland, 2007), in more naturalistic settings this repetition can be recurrent. Speakers who have agreed on a specific reference or 'label' for a given item will usually repeat this same label throughout a conversation, as changing it would be assumed to indicate a change in meaning (i.e. if a speaker has referred to a figure as "*the slender one*", suddenly referring to it as "*tall and thin*" would cause conflict in the hearer, who might need to look at the context to check whether a different figure is being referred to, as in Barr & Keysar, 2002; Metzger & Brennan, 2003; Senay & Keysar, 2009; Shintel & Keysar, 2007; for the general 'Principle of Contrast' see E. Clark, 1987). Through this repetition, a convention will be formed: the reference that was used will be maintained not only because it was primed by its previous use(s), but because it has been established as a successful way of dealing with the need of referring to that item among those speakers.

In this review, I will investigate the literature on the emergence and establishment of referential conventions in dialogue (with occasional references to both monologue and multi-party conversations), considering the influences of context and participant role, and the possibility of change of an established convention. While it will be explicitly stated when an element pertains to one of the above mentioned theories exclusively, the chapter's focus combines elements of both approaches, as it deals with dialogue 'anomalies' that are, by definition, not considered by the alignment model (Pickering & Garrod, 2004).

## **2.2 The production of references in dialogue: Historical and ahistorical factors**

Dialogue is the most natural setting for language use. Usually, interlocutors engaged in dialogue share a physical and social setting, which is used as background to aid in the interpretation of the meaning of words and phrases being uttered. Moreover, non-concurrent context can also be brought to the production and comprehension processes, if both speakers knowingly share a past history of interaction or a group membership (cultural co-presence or community membership; Clark & Marshall, 1981). The following section will consider both the factors that influence the initial production of references (particularly salience in context and level of specification), and the historical factors that influence whether a reference is repeated or discarded over time.

### *Ahistorical factors*

When a speaker in dialogue produces a reference, she is looking for an effective and efficient way to communicate a meaning (even though she might not be aware of this search). She will aim at a reference that is at the same time easy enough to produce and specific enough as to avoid the effort of having to engage in additional explanations or repetitions to repair a misunderstanding (Grice, 1975; Schegloff, Jefferson, & Sacks, 1977). To do this, she will consider both purely linguistic factors, like lexical availability (derived from the word's frequency, syntactic position, and previous use, among other factors, see Anderson, Garrod, & Sanford, 1983; Bock &



Warren, 1985; Ferreira & Dell, 2000), and more contextual factors, like informativeness in context (discriminateness between the target and other possible referents; Brennan & Clark, 1996), and perceptual salience (salience of possible referents in the immediate context; Tarenskeen, Broersma, & Geurts, 2015; Vogels, Krahmer, & Maes, 2013).

As Tarenskeen et al. (2015) have suggested, the basic idea behind salience is intuitive: speakers tend to select attributes of an object or situation according to the degree to which their attention is attracted by them. Communicatively, this is a sensible strategy, as it is likely that those attributes will also attract the attention of their interlocutor (however this need not be a conscious aim). In this sense, a speaker choosing which attribute(s) to highlight when referencing an object (e.g. *“the big square bag”* vs *“the fake-leather bag”* vs *“the square white bag”*) is, consciously or not, assuming that their focus can be shared with little or no effort by their interlocutor, which would make the use of those attributes relevant in the process of correctly identifying the referent (Clark, Schreuder, & Buttrick, 1983). However, the inclusion of salient features in an item’s reference can sometimes override the aim for a sufficient but not overly informative description, as the most salient attributes of the item are not necessarily the ones that are enough to distinguish it successfully in its context (Brennan & Clark, 1996).

Salience is also related to the surprise value of the object or attribute considering both its context and its known features (Itti & Baldi, 2009). In this sense, a different object in a context of similar ones would have high salience, as well as an uncommon feature in a common object (a square plate, or a blue rose). Salient entities or attributes are also assumed to be more accessible in memory than less salient ones (Vogels et al., 2013), which improves the likelihood of their mention contributing to the successful identification of the referent. Vogels et al. (2013) manipulated both the linguistic salience of referents (by having referents in either subject or object position in a phrase) and their visual salience in accompanying images (by having referents in either foreground or background of a picture), and found that visual salience, but not linguistic salience, had a reliable influence on speaker references, with visually salient entities chosen as main referents significantly more than less salient entities.

Overall, the evidence seems to confirm that contextual salience is perceived as a shared starting point that modulates the likelihood of certain aspects of the scene to be chosen in the construction of a reference. This process, where the speaker uses their own salience judgment as a basis for the attribution of a similar judgment to their interlocutor, can be linked to epistemic egocentrism, the use of the individual's own mental state as a template for their modeling of others' minds (Keysar, Lin, & Barr, 2003; McClung & Reicher, 2018).

However, salience on its own cannot explain the inclusion or exclusion of information from a reference, as the communicative needs of the speakers might be more or less detached from the attentional landscape of their context of interaction. If the speaker needs the addressee to identify a specific referent, she has three alternatives in terms of information inclusion: speakers might underspecify their referential descriptions, producing potentially ambiguous references, or descriptions that do not uniquely specify a referent among a set of options (Ferreira, Slevc, & Rogers, 2005). Alternatively, speakers might overspecify their references, including more information than is needed to uniquely identify the referent (but see Rubio-Fernández, 2016, on why an *efficiency*-based analysis might be better than standard pragmatics to discuss reference production). Naturally, speakers might also include exactly enough information as to allow the addressee to identify the referent (but even this might be sub-optimal in some contexts, see Clarke, Elsner, & Rohde, 2013).

In natural language use, overspecification is much more common than underspecification (Ferreira et al., 2005; Sedivy, 2003; van Deemter, Gatt, van Gompel, & Krahmer, 2012), as the latter is more likely to increase the overall effort of the conversation by requiring repair sequences or additional turns. Overspecification, on the other hand, is particularly abundant with respect to intrinsic properties of objects such as colour or number, which are generally assumed to be least likely to cause processing problems (Brown-Schmidt & Konopka, 2011). Sedivy (2005) showed that speakers do not always assume contrastive interpretations of modifiers (particularly in the case of colour adjectives), relying instead on communicatively based expectations, when interpreting references to objects in an elicited-production and comprehension task. Sedivy's results suggest overspecified references can be interpreted as simply as basic-level descriptions,

without invoking additional objects that would justify the inclusion of additional referent properties (Engelhardt, Bailey, & Ferreira, 2006).

Even though overspecification is frequently produced in cognitively undemanding contexts, it is more frequent when the importance of the task is high, and the overspecified attributes could potentially help referent identification, suggesting goal-directed use (Arts, Maes, Noordman, & Jansen, 2011; note however that this is not indicative of conscious control, see Wardlow Lane, Groisman, & Ferreira, 2006). Other factors that could contribute to the production of overspecified references are incremental production, division of attention, failures in the determination of redundancy, and expected identification difficulty (Arts et al., 2011).

The overspecification of colour in descriptions where colour is not needed to identify a referent is particularly common in natural language (Tarenskeen et al., 2015). In spontaneous mention, Sedivy (1999) found that more than half of references to common objects included colour when this was not needed for identification. According to Rubio-Fernández (2016), colour is likely overspecified due to its saliency (making it more likely to grab both speakers' and listeners' attention) and contextual contrast properties (other elements in context are likely to have different colours, with this feature therefore acting as potential discriminator). Even in studies that showed a detrimental effect of overspecification, colour was found to have a much smaller negative effect on referent identification than size overspecification, which caused a negative effect twice as big (Engelhardt, Demiral, & Ferreira, 2011). One side of this issue pertains to the context-dependency of features used in overspecification: while size is usually expressed as a scalar adjective (e.g. "*the tall glass*" in comparison to either other co-present glass, or an implicit standard of glass sizes, or "the smallest size available"), therefore requiring some form of co-present or implicit comparison standard, colour is generally expressed in absolute terms, and is therefore independent of other elements in context. As such, it should be possible to overspecify colour by default without risking misleading the hearer into assuming a comparison class.

The question of whether overspecification is beneficial, neutral, or detrimental for comprehension has been hotly debated in the last decade. While Engelhardt et al. (2011) found a comprehension delay related to the inclusion of overspecific

adjectives (though this effect was likely partially driven by predictability), others have found processing benefits. Both Paraboni & van Deemter (2014) and Clarke et al. (2013) showed how overspecification can benefit the hearer if it facilitates a spatial search, by including additional information that can help the hearer narrow down the search space. Paraboni, van Deemter, & Masthoff (2007) showed that speakers are more likely to produce the same types of overspecification that hearers find beneficial, thereby closing the link between production and comprehension. In spatial search tasks of different difficulty levels, they found speakers produced more logically overspecified spatial descriptions when the spatial search task was harder for the addressee, who would benefit from the additional information in the search (e.g. a speaker would produce “*picture 2 in part B*” instead of “*picture 2*” alone when there was only one picture 2, if the search space included more subsections). In terms of listeners’ expectations, Engelhardt et al. (2006) found that listeners do not consider overspecified descriptions to be any worse than appropriate descriptions, suggesting there are no significant additional processing costs associated to their comprehension (however, they used a task where subjects had to judge the effectiveness of an instruction to move an object in producing a desired outcome, which might give different results compared to an interactive task where the instructions were actually used in communication).

Taken together, most studies agree that context is crucial in determining whether speakers are likely to overspecify a reference or not: if the context is cluttered or items are hard to identify, speakers are more likely to include additional information (Clarke et al., 2013). On the other side, in extremely simple contexts (such as two-figure displays of geometrical shapes) overspecification might cause comprehension delays as speakers do not naturally overspecify in these contexts (Rubio-Fernández, under review), which might prompt hearers to attribute additional meaning to overspecific references. Overall, most researchers agree overspecification is not usually detrimental for the listener and might be even beneficial in contexts where the overspecific features are not present in other competitor objects (Rubio-Fernández, 2016; Paraboni & van Deemter, 2013).

Contextual and perceptual factors affect reference choice in multiple ways. Not only the perceptual properties of the referent and the context might guide speaker’s word choice, but they also influence their conceptualisation of the object, making it more

likely that they will choose one perspective over others if specific properties are more salient (Anderson & Garrod, 1987). This is particularly true when there are multiple possible perspectives on an object, such as in the case of the maze game (Garrod, Anderson, & Sanford, 1984; Garrod & Anderson, 1987; Garrod & Doherty, 1994): mazes can be conceptualised as sets of rows and columns, paths from one point to another, figurative structures, etc., and the reference choices of the speaker facing a specific maze will reflect the conceptualisation they are using to interpret that maze.

When a speaker chooses words to refer to an object, they are portraying a specific perspective on the object, a conceptualisation that will communicate to the hearer which aspect of the object is to be observed, or how is the object to be understood (E. V. Clark, 1997; Van Der Wege, 2009). In that sense, the relationship between the reference and the object is mediated by this conceptualisation, such that when a speaker uses the reference to communicate a specific perspective on the object to a hearer and the hearer accepts it, it is the perspective that is being accepted by the hearer. Indeed, perspective choices reflected in references affect how people remember objects (Carmichael, Hogan, & Walter, 1932). In a communication study, Wilkes-Gibbs (1995) showed that people are slower to identify an object if they were primed to refer to it with references reflecting different perspectives (such as “*the flagman*” or “*the angel*” to refer to the same tangram figure). The choice of repeating a reference or not, therefore, implies a decision with respect to the acceptance of the perspective the speaker is putting forward.

### *Historical factors*

When a speaker introduces a referent to a conversation, their choice of referring expression will create a link between the chosen linguistic expression and the referent that will stand in the interlocutors’ memory as part of their mental model of the discourse, acting as an attractor for subsequent references to the same or similar items (Arnold, 2008). This first reference, if accepted by the interlocutor, constitutes a *precedent* for subsequent interactions, that is, an alternative that has been established as valid by its successful use in a previous interaction. As a link between the expression and the referent has been created and accepted (explicitly or tacitly) by the interlocutor, both the original speaker and listener will expect that

the same referring expression should be reused whenever the same referent is brought to the conversation, as this precedent stands out as a salient option against other unattested alternatives (Kronmüller & Barr, 2015; Metzling & Brennan, 2003).

A recent review of the work on precedents and the impact of a change in reference use (Kronmüller & Barr, 2015) has shown that there is strong evidence that listeners expect references to be maintained (that is, that the same linguistic expressions will be used to refer to the same referents), which is consistent with most dialogue models. Recent studies have also found comprehension is facilitated if speakers are consistent, that is, if they use the same references for the same objects over time (Horton & Slaten, 2012; Kronmüller & Barr, 2015).

The most problematic aspects of the discussion pertain to the effects of speaker identity in maintaining or breaking a precedent. While most studies agree there is an advantage for the listener if the same references are used for the same referents, there seem to be additional effects associated with the identity of the speaker. Comprehension appears to be facilitated if the same speaker maintains the same references, compared to a different speaker using the same references, though the strength and duration of this effect is not clear, appearing to be brief and concentrated on the earliest moments of comprehension (Kronmüller & Barr, 2015). The breaking of a precedent, on the other hand, while costly to the listener, appears to be less costly if produced by a new speaker, versus the same speaker that had previously used the reference (a 'different speaker advantage for broken precedents', Metzling & Brennan, 2003).

However, Barr & Keysar (2002) have suggested that the main driver behind the use of precedents is not that they are mutually known, but that they are egocentrically available. According to their model, speakers rely on a fast and automatic egocentric heuristic in the use of information to process linguistic input, coupled with a slower, optional 'common ground' system that specifies (if needed) the perspective of their interlocutor. Much of the reported benefit of maintaining linguistic referents is assumed to derive from memory-related (egocentric) processing, such that any item that has been previously used (by any speaker) presents a processing advantage over a new item. This idea assumes speakers will rely on any available information, independent of source, to constrain and ease the effort of language understanding.

However, this benefit is assumed to be modulated by a *generation effect*, according to which the speaker producing the linguistic item may have better memory for what has been said in conversation, compared to listeners, since they 'generated' those references (Slamecka & Graf, 1978; Knutsen & Le Bigot, 2014; Knutsen, Le Bigot, & Ros, 2017).

Overall, it can be assumed that, if a reference has already been used successfully in conversation, the same item will be reused by those involved in the conversation. The degree to which speakers adhere to this precedent has been related to the role they performed in the dialogue, with more precedent use being associated to a more central role in the dialogue (as a speaker or addressee, as opposed to side-participant or overhearer) (Branigan et al., 2007). However, precedents are also strengthened by the frequency of their use, with a higher frequency of use in a task being associated to a stronger precedent (Garrod & Doherty, 1994; Brennan & Clark, 1996). Frequency of use has been described as a direct influence on ease of activation of a representation in memory, both at a local level (i.e. the frequency of use of an expression in a conversation, Brennan & Clark, 1996) and at a language level (Rayner, 1998). As such, the likelihood of use of a precedent should be directly related to the frequency of its past use, establishing a cumulative measure of precedent use.

While most of the factors reviewed so far pertain to the domain of individual processing, the needs of the addressee are also considered in reference production. In an image-sorting experiment, Brennan and Clark (1996) showed that participants were more likely to retain their previously used references if they were talking to the same partner as before, compared to talking to a new partner, even when these references became overspecific due to a change in context. If talking to a new partner, speakers would either reduce their descriptions to a basic-level label, removing the overspecific information, or expand their references including more information as before. The same study showed that the likelihood of maintaining the same reference that had been previously used was related to the number of times the reference had been used with that partner.

In order to produce a communicatively adequate reference, the speaker needs to consider what the addressee is likely to know (information that should be considered

*common ground* between the interlocutors), and what constitutes new information. To do this, she needs to establish *global* and *local* assumptions about the addressee, derived respectively from the general status of the addressee (cultural membership, age, gender, etc.) and his specific knowledge about the conversational topic and context (Arnold, 2008). While many studies have shown that speakers adapt their linguistic production to the assumed needs of their partners (Isaacs and Clark, 1987; Fussell & Krauss, 1992), there is a logical limit to this process, derived from the cognitive load of the additional assumptions required to create a detailed representation of the addressee's knowledge (Bard & Aylett, 2005). In the next section we will analyse the establishment of common ground and the constraints on its use in reference production.

### **2.3 Common ground**

The term *common ground* is used to describe the mutual knowledge that is established between speakers during a conversation, which they update as the conversation develops (Stalnaker, 1978; Clark & Carlson, 1981; Clark, 1996). In its broader definition, it includes all their common knowledge, beliefs, and suppositions (plus their awareness of their commonality), even if these do not immediately enter the linguistic interaction (Clark, 1996). The derived concept of *grounding* is used to refer to the process by which speakers in interaction affirm their mutual acceptance of stated or implied knowledge, usually through various forms of acknowledgement, backchannels, and repair (Clark & Brennan, 1991; Fusaroli, Tylén, Garly, Steensig, Christiansen, & Dingemanse, 2017). The idea of grounding also implies that both interlocutors believe they have understood what the contributor meant, and that it can be used from then on in the conversation as part of the common ground between them (Clark & Schaefer, 1989).

The idea that common ground plays a role in comprehension is generally accepted, however, there is no clear consensus on how this knowledge is represented, used, and stored in conversation. Specifically, there are conflicting accounts of how comprehenders use their knowledge of their common ground with the speaker in interpreting their utterances (Keysar, Barr, Balin, & Paek, 1998). As experimental work has shown that listeners will consider information explicitly not known to the



speaker when interpreting the speaker's expressions (Keysar, Barr, Balin, & Brauner, 2000; Keysar et al., 2003), some accounts have suggested that the use of common ground in interpretation comes after (in processing and in time) an initial, purely linguistic, stage (Kronmüller, Noveck, Rivera, Jaume-Guazzini, & Barr, 2017); cf. Hanna, Tanenhaus & Trueswell, 2003). Keysar and colleagues (in their Perspective Adjustment model, also referred to as *anchoring and adjustment* model; 1998, 2000) suggest egocentric linguistic associations constitute the main tool for the determination of possible interpretations for a reference, with common ground accessed only if needed, in a slower and more complex second stage of processing.

Barr & Keysar (2002) showed experimentally that the benefit of common ground is primarily related to the availability of the previously used linguistic items (precedents) for the speaker, and not to an association with a specific interlocutor (mutual knowledge or conceptual pacts). However, their crucial test involved speakers inhibiting a precedent when faced with a new interlocutor. Metzling & Brennan (2003) argue that speakers do not inhibit a precedent just because they are speaking to a new partner, and show that speakers are aware of the entrained terms they are maintaining with each partner, reacting if these precedents are broken (i.e. an old partner using a new term for a previously named referent). Kronmüller and colleagues (2017) have recently suggested that the use of common ground in comprehension is not activated immediately, instead appearing after an initial 'speaker-independent' processing in which the benefit of a precedent is not linked to any particular speaker. In a similar line, Horton & Keysar (1996) suggested that considering the perspective of the addressee only occurs after an 'egocentric' first stage, showing that speakers under time pressure did not consider their partner's perspective, but speakers who had more time did.

A more nuanced view is held by Brown-Schmidt and Hanna (2011), who argue that common ground is one of many partial constraints on comprehension that are resolved by the hearer through weighing different sources of evidence, and that those weights can vary according to context and speakers' goals (Brown-Schmidt, Yoon, & Ryskin, 2015). In the same line, Yoon, Koh, & Brown-Schmidt (2012) suggested speakers consider their addressee's knowledge more strongly when they have to request something from them, as opposed to merely informing them of something.

A memory-based account, put forward by Horton and Gerrig (2005, 2016), suggests most of the processing usually attributed to common ground demands is better explained by memory-general processing of event information (episodic memory), which registers the identity of the speakers along with all other relevant information about the event. According to the authors, the demands of the traditional common ground literature (diary-like representations, co-presence evaluation for each speaker-addressee-referent combination, etc.) would be too cognitively costly to constitute the default method of retrieving referential information. In their view, the identity of dialogue partners is associated to the reference expressions used with them, such that the activation of those expressions will be stronger when interacting with the same partner than with a different partner. Thus, it is more likely that the same references will be reused if the same partner is re-encountered, as these expressions become more readily accessible in memory through resonance (Ratcliff, 1978; Knutsen et al., 2017). Importantly, the strength of the activation would be modulated by use, with more use being associated to a stronger posterior activation.

Even though most of the literature has focused on the definition of a default processing mechanism that could explain why some processes appear to be fast and automatic while others are slow and 'costly', more recent work has questioned this view. Dale and colleagues (Dale et al., 2018) argue that perspective-taking requires the continuous integration of information acting at different timescales, with no 'default' mechanism that could explain the final outcome. While some processes might be faster than others, integration is performed on the basis of relevance in context, meaning 'slow' processes might be more cognitively pervasive than 'fast' ones. In a similar tone, Degen and Tanenhaus (2015, 2016) argue for a situated approach, where the integration of information is constrained (and thus facilitated) by the context and the demands of the task.

### *Referential reduction*

When a reference is first used in conversation, the addressee is expected to signal whether it is accepted or not. According to Clark and colleagues (Clark & Carlson, 1982; Clark and Schaefer, 1987, 1989), the *Principle of Responsibility* implies that

continuing the dialogue can be assumed to signal that all past references are accepted as such by both interlocutors, as it would be the responsibility of whoever encountered a problem to indicate the occurrence of a misunderstanding. As dialogue is inherently repetitive (Tannen, 2007), particularly in terms of entities, places, or activities (Healey, Purver, & Howes, 2014), references used in conversation tend to be reused, leading to a process of simplification or reduction.

The repeated use of a label in referring to an item usually entails a dynamic process of *referential reduction* (Krauss & Weinheimer, 1964, 1966; Clark & Wilkes-Gibbs, 1986; Metzing & Brennan, 2003), where a longer initial description (usually an indefinite descriptive phrase) is reduced to its main constituents (usually a definite noun or noun phrase). This phenomenon has been traditionally analysed as a consequence of the interactive grounding process between speakers, which entails a move from the reference as a description (pointing to a meaning in the world), to the reference as a sign that points to its own past use, that is shared between speakers (Garrod, Fay, Lee, Oberlander, & Macleod, 2007). As such, this reference doesn't require the amount of detail that would be needed to retrieve the referred item among the alternative options in a visual scene, but only enough to identify the previously used description that was successful between the speakers to refer to this item. Therefore, the speaker can 'drop' the elements in the reference that were useful when it was pointing to the context, such as details or modifiers, but which are no longer required when it refers to its previous use: a successful use of "*the white leather bag with the golden handle that's over the table*" to refer to a specific bag can become "*the white bag*" on subsequent references.

This dropping of hedges and other modifiers is taken to mark the conceptualisation as accepted, as hedges and modifiers communicate uncertainty about the definition that is being used (e.g. "*a kinda reddish lab*" becomes "*the red lab*" once the speaker can be certain the reference was understood) (Brennan & Clark, 1996; Metzing & Brennan, 2003). As a whole, references tend to become shorter and more efficient as they are used by an interacting pair, compared to a speaker talking to a recorder or to an imagined partner (Van Der Wege, 2009).

The referential reduction process has been characterised as a move from an indefinite to a definite reference (Clark & Wilkes-Gibbs, 1986), or from a description

to a label. But what is it exactly that is being 'reduced' in this process? Using an online tangram game, Hawkins, Frank, and Goodman (2017) showed that function words were eliminated more frequently than content words. In their description of the process, references in an initial stage are characterised by an uncertainty on the part of the speakers that translates into multiple partially redundant phrases, which are dropped as the game progresses and the ambiguities are solved. The evidence points to a move towards the minimum-length label that is enough to indicate to the interlocutor that a previous reference is being activated.

Similarly, referential expressions are simplified with repeated mention as their antecedents become more easily accessible in memory (Ariel, 1990), for example going from a full nominal phrase "*my friend's big brother*" to a pronoun "*him*". According to Bard and Aylett (2005), egocentric processing explains most of this simplification. In their analysis of the dialogues in the HCRC Map Task Corpus, they showed that neither listener's co-presence nor feedback affected the rate at which speakers simplified their referential expressions, which were simplified mainly as a result of repeated mention.

In graphical communication, a similar process explains the evolution of graphical symbols from icons that get their attributed meaning from their similarity to the represented objects, to symbols with a conventional meaning. As a referential relationship is established between an icon and the meaning it stands for, the locus of information shifts from the sign as a direct representation of the meaning towards the users' memory of the past usages of the sign, that is, towards the meaning-symbol relationship they have grounded through interaction (Garrod et al., 2007). In this process, the sign's iconicity gives way to its abstraction as a symbol of itself, requiring only enough information as to signal its status as convention pertaining to the interlocutors' common history. The similarities and differences between these simplification processes in linguistic and non-linguistic communication will be further discussed in section 2.7.

## 2.4 Interaction and participant role in conversation

So far we have described conversation as an experience of two participants, each of whom performs the role of speaker or hearer interchangeably throughout the dialogue. However, dialogues frequently include other participants, who might be actively involved in the conversation, or only listening to it without taking an active part. Moreover, as participant role is defined according to the speech act that is being performed, a conversation can have the same individuals in different participant roles at different points: a speaker can be an active interlocutor for one topic of the conversation, but become a side-participant when a new interlocutor introduces a different topic. Clark and colleagues (Clark & Carlson, 1982; Clark & Schaefer, 1987) proposed a classification that is based primarily on accountability: *side-participants*, who are not addressed by the speaker, are accepted as participants in the conversation and have rights and responsibilities as such, like being expected to share the common ground that is established in the conversation; while *overhearers* are not considered to be participants in the conversation even if they have access to it, and are not expected to respond to it in any way nor accumulate common ground as a participant would do. Further, *eavesdroppers* are non-participants whose presence or capacity for listening in is not acknowledged by the participants of the conversation.

Participants in a dialogue adapt their expectations of what can or cannot be comprehended by an individual according to the role she has played in the dialogue: while side-participants are expected to share most of the common ground built during a dialogue, overhearers are not (Clark & Carlson, 1982; Wilkes-Gibbs & Clark, 1992). According to Wilkes-Gibbs and Clark (1992), speakers assume side-participants who are acknowledged as participants in the conversation (even if not directly addressed) should share common ground with them as much as direct participants in the conversation, in opposition to overhearers who are not expected to share this common ground even if they had access to the same conversation. However, this capacity has been questioned (Branigan, 2006; Branigan et al., 2007), as side-participants have limited options to either confirm or disconfirm their understanding, and might show differences in their accumulated discourse record.

The evidence suggests differences between side-participants and overhearers are to be expected in comprehension tasks, but that this would be mediated by access to referents and other contextual elements in the conversation. In a task where speakers described tangram figures to hearers accompanied by side-participants or overhearers with different levels of access to the main interaction, Wilkes-Gibbs and Clark (1992) found that overhearers took longer to identify the referents of descriptive phrases, compared to side-participants, and produced more identification errors. Similarly, Schober and Clark (1989) found that overhearers who just listened to the recorded interaction made more errors than addressees in identifying referents in a tangram identification task. While no research so far has identified a direct cause for these comprehension failures, the acknowledgement side-participants get from speakers might involve some form of audience design that includes them as listeners, even if they are not addressed directly.

With respect to alignment, side-participants and overhearers have been also shown to behave differently. Working with syntactic alignment, Branigan and colleagues (Branigan, Pickering, MacLean, & Cleland, 2007) showed that both addressees and side-participants align with the speaker in conversation, but addressees align significantly more than side-participants. Behnel, Cummins, Sichelschmidt, and De Ruyter (2013) found a significant, but low, alignment effect in overhearers who were not involved in the task the main participants were performing, and a much higher effect in overhearers who were involved in the task by having access to the same perceptual context the main participants were exposed to. Taken together, these results suggest alignment is modulated by participant role and involvement in the interaction, as the same linguistic exposure bears different alignment and comprehension consequences for participants in the dialogue than for non-participants.

### *Dialogue vs monologue*

Even though we assume dialogue to be the most natural setting for language use, and monologue to be almost entirely absent in everyday activities, there are some advantages to the study of the specificities of language use in dialogue vs monologue. Particularly, the trajectories of language in use in a dialogue, in terms of conservation and change of referential choices mediated by the interaction of

speakers, might be quite different from the trajectories of language in use without interaction.

Two features of language in use that allow for comparison between dialogues and monologues have been studied: the referential reduction process (described in the previous sub-section), and the comprehension benefit to third parties in language extracted from dialogues vs monologues. Hupet and Chantraine (1992) employed a referential task similar to Clark and Wilkes-Gibbs (1986) with participants describing tangram figures to a future participant in two conditions: a same listener condition, where they believed all their descriptions would be delivered to the same person, and a different listener condition, where they believed their descriptions would be received by different people. Afterwards, they compared these results to the outcome of the same task performed by interacting pairs (Hupet, Seron, & Chantraine, 1991). Their results showed speakers in the Same Listener condition used fewer definite references and labels than speakers in dialogue, and did not reduce the length of their descriptions, while speakers in the Different Listener condition used almost no definite references or labels; in fact, speakers in this condition used more words to describe each figure as they advanced in the game, and required at least the same time to complete the task in the last rounds as they needed in the first rounds.

Fox Tree (1999; see also Rogers, Fay, & Maybery, 2013) suggested that there is a comprehension benefit that stems from dialogue, as opposed to monologue, such that an overhearer would be better able to identify a referent from a dialogue rather than a monologue, even if the monologue was more detailed in its descriptions. However, it was not clear whether this benefit stemmed from a property of interaction (feedback, pace, etc.), or from the higher number of perspectives that is usually associated with natural dialogue. To clarify this point, Fox Tree and Mayer (2008) designed a task that was balanced between monologues and dialogues in the number of perspectives each description presented, and found that the comprehension benefit was attributable to the number of perspectives alone, across monologue and dialogue conditions. However, a later study by Branigan, Catchpole, and Pickering (2011) showed that the grounding of a perspective (whether it has been accepted by both speaker and addressee) constitutes a more important factor in the comprehensibility of the message for third parties. Using a similar tangram

description task, Branigan and colleagues found that overhearers were more accurate in their identification of the figures when the description they heard had been produced in a dialogue setting, even when the description itself only included one perspective (descriptions from later rounds, in which a perspective has already been grounded). Taken together, these results highlight the relevance of the interactive grounding processes in the establishment of a conventional reference that is useful not only to the interlocutors but also (at least potentially) to others, in a process that usually includes negotiating different perspectives, but that might not depend on this factor.

## 2.5 Conventions and change

Language is conventional in a broad sense, as the relationship between signals and meanings in any given language is largely arbitrary (though the impact of iconicity in language evolution, acquisition, and as a general property of language is still under discussion, see Dingemanse, Blasi, Lupyán, Christiansen, & Monaghan, 2015; Imai & Kita, 2014; Perniss, Thompson, & Vigliocco, 2010). As a consequence of this arbitrariness, pairings of signals and meanings need to be learned by the speakers before they are able to use a language communicatively. But linguistic conventions also develop in a narrower sense through the use of language in interaction, in specific conversations, or among particular groups of speakers. These linguistic conventions are defined, in Lewis' terms (1969/2002; see also Croft, 2013), as stable coordination equilibria that people create to solve recurrent coordination problems: if two speakers need to refer repeatedly to some aspect of the context, they will develop a conventional reference that will be considered as the default option to deal with this need.

Garrod and Pickering (2013, see also Garrod, 2011) use the term *routinisation* to describe the process by which a specific word(s) and meaning combination becomes conventional in a pairwise conversation or group of people. This routinisation process eases the processing effort of the speakers by eliminating the need to evaluate alternative expressions to communicate the same meaning (*“the pointy bit”* of a figure could also be referred to as *“the sharp end”*, etc.), or alternative meanings that are linked to the same expression (if we both use *“the teacher”* to



refer to a specific individual in a conversation, we don't have to think about all the other "teachers" that we know). In this sense, the process shifts the locus of information of the chosen expression, from a meaning in the world, to the speakers' memory of its past use, making it at the same time easier to process for the speakers and more opaque to outsiders (Garrod et al., 2007).

Conventions or routines are generated frequently in conversations of some length, as the repetition of topics and referents is common in natural language (Tannen, 2007; Garrod & Pickering, 2013). As such, they might or might not outlive the conversation, depending on the likelihood of repetition of a topic for that specific pair of speakers. An expression repeatedly used between speakers might become a convention or routine, as its surface form becomes lexicalised for the speakers through the setting down of new memory traces that link the expression with that particular meaning (Garrod & Pickering, 2013). As such, the strength of the convention will vary according to its frequency of use with that same meaning by those speakers, showing that, even though there is an automatic, mechanistic aspect of priming, usage modulates its impact (Oben & Brone, 2016; Pickering & Garrod, 2013; see also Brennan & Clark, 1996). This cumulative effect of priming in the formation of conventions or routines can explain the link between the dialogue level and language change, which will be discussed in the next subsection.

Even though use modulates the strength of a convention between pairs of speakers, in a broader sense the strength of a convention depends on the existence of a community of speakers that can be expected to maintain it even if a specific pairwise interaction does not abide by it. In this sense, two speakers can only sustain a *local* convention, which is susceptible to being replaced, as the use of any other alternative undermines the certainty of the speakers in its status. As Lewis (1969/2002) stated, each instance of a regular behaviour that solves a common coordination problem adds up to the belief that that is what members of this community or group do when they face that coordination problem. As such, a convention represents the standard that members of a community can look to, knowing that all other members expect that convention to be used. In that sense, it relies on a community level representation, as the individuals base their choice on a generalisation from their individual experiences to the entire community (Barr, 2004).

However, experimental work has shown that explicit community-level knowledge or assumptions are not necessary for the emergence of conventions (in a natural language setting, Garrod & Doherty, 1994; in larger populations through web-based communication, Centola & Baronchelli, 2015). Work on simulation models has also shown that conventional behaviour can emerge without explicit community-level knowledge or central control, as an emergent outcome of dyadic interactions repeated over time (Baronchelli, Felici, Loreto, Caglioti, & Steels, 2006; Barr, 2004). In these models, the information each agent holds is updated on the basis of the success or failure of its current interaction, and then used to predict what other agents will do when facing the same situation (from an initial random variation, the result of the interactions adds weight to one or other alternative behaviours). The community-level convention would be sustained, in this framework, by a sequence of dyadic interactions, without the need for explicit or global coordination.

In real life communication, though, speakers are usually aware of the identity of their interlocutor, and have enough memory to maintain a rough idea of how common a given convention is, as a result of their interactions with different members of the community (though their estimates are usually egocentrically biased, see Krauss & Fussell, 1991). Pairwise conventions are therefore linked to the identity of the interlocutor, as we saw in section 2.3. As such, speakers can sustain different conventions to refer to the same meaning when interacting with different individuals/groups, such as a speaker using different jargon to refer to her sports club when with her teammates vs. when with her family. This flexibility serves as a reminder of the arbitrary nature of conventions, as the same meaning can be communicated through completely different signals, and even an established convention can be abandoned for a new alternative in the right context.

### *Adaptation and change*

Although most research on reference in dialogue deals with the ways in which speakers repeat each other and/or converge into unique, pairwise choices, there are several reasons and scenarios in which speakers will change the reference they were previously using. Changes in context, linguistic repetition, or role differentiation are some of the common causes behind changes in conventional pairwise

references. Brennan and Clark (1996) have suggested that references in pairwise conversation are always provisional, as they can be changed or abandoned in the right context even if a precedent was established, since each instance of a reference acts as precedent for the next. This argument can be linked to the observation by Lewis (1969/2002) that a convention between two people is always weaker than a group convention, as there is no possible appeal to the group choice that could stabilise the conventional option when one individual changes their choice.

One side of this issue pertains to the question of perspective or topic differentiation. One of the critiques of alignment as a main explanation for sense-making in conversation is that of Healey and colleagues (Healey et al., 2014; Healey, Mills, & Eshghi, 2016), who argue that the repetition usually associated with alignment does not explain the 'forward momentum' found regularly in conversation, as speakers need to diverge and introduce new perspectives and topics for a dialogue to be meaningful. In their analysis of syntactic structures in corpus dialogues, they found speakers repeat each other less than what could be expected by chance, diverging from one another in their use of syntactic constructions and not even repeating their own structures. Even though their results clash with most experimental work in syntactic alignment (Branigan et al. 2007; 2011; etc.), their use of natural conversations (where there is no predefined topic and which might be varied in length, aim, degree of acquaintance of the speakers, etc.) might open a window into the interplay of alignment and other forces in natural conversation, particularly complementarity and divergence.

In some situations, complementary roles are required to maintain the flow of conversation (either by task demands, or naturally in the course of a dialogue), shifting the behaviour of the interlocutors away from repeating a conventional choice. Coco, Dale, and Keller (2018) have suggested that alignment on its own should not be considered a proxy for effective communication, insofar as the demands of the communicative task might require different behaviours from the interlocutors. Experimental studies have found that indiscriminate local alignment was linked to lower performance on a collaborative task, as opposed to selective alignment of task-relevant expressions (Fusaroli et al., 2012). Similarly, Garrod and Clark (1993) found that children working on a maze game task would converge on a

superficially similar description scheme without a common interpretation of it (e.g. they would both use *“the top line”* to refer to a position, but one would mean the top overall row, while the other would mean the first complete row), with younger children (7-8 years) showing more repetition of their partners' expressions but less task success than older children (11-12 years).

The other side of the issue is change in reference choice. Usually, these changes are produced to accommodate to a change in context, to respond to varying communicative needs, or to adapt to the requirements of the task. When a conversational situation changes, established conventions may no longer be appropriate, and successful coordination might require speakers to adapt their referential behaviour dynamically to better fit the new situation. Speakers can and quite frequently do change the way that they refer to referents in natural conversation (Healey, 2008), even though this is assumed to be costly for the speaker (and the listener, depending on the context) (Brennan & Clark, 1996; Kronmüller & Barr, 2015; Metzinger & Brennan, 2003).

Ibarra and Tanenhaus (2016) used two different tasks employing the same items in order to test the flexibility of speakers in their use of conventional references. In the first task, participants described different unfamiliar objects, establishing common references to identify each object (e.g. a long, flat object would be called *“the wrench”*); while in the second task, the same objects would be employed in the construction of a structure in which each object had to be placed in a particular position (e.g. the same object would now be used as the leg of an animal). Their participants would abruptly abandon their entrained references in order to switch to task-relevant references (e.g. moving from *“the wrench”* to *“the leg”*), usually without negotiation, and without any explicit instruction to do so. The authors discuss these results highlighting the role of task goals in reference choice, suggesting even a grounded conceptual pact can be abandoned if the task requirements change. Moreover, their results suggest speakers can flexibly use and maintain in their memory different conceptualisations for the same items.

Reference change might also stem from production needs, as in lexical differentiation. Lexical differentiation explains why speakers use a different reference when naming a second object that belongs to the same category as a first

object, even if the two are not in the same context and therefore not competing for the same label (Van Der Wege, 2009; Yoon & Brown-Schmidt, 2014): If a speaker has called an object *“the shirt”* when shown in a context of other, unrelated images (a car, a plant, etc.), they will call a different shirt, shown afterwards in a new unrelated context, *“the striped shirt”* or *“the blouse”*, even though the same basic label *“shirt”* could be applied successfully in this second scenario. Van Der Wege (borrowing from Clark and Clark, 1979) call this phenomenon “pre-emption by similar form”: speakers will avoid giving two different items the same label, even in two different scenarios (when it would not be contextually maladaptive), as the first use of the label establishes a link with the item that prevents any other item being associated to the same label in the conversation. Similarly, Garrod and Anderson (1987, see also Garrod & Doherty, 1994) discussed how participants in a maze game task will change the names they use to identify reference points when switching to a different counting convention, e.g. moving from *“top line”* to *“top row”*. These changes in reference choice show that references are associated to specific conceptualisations, and that the link between reference and conceptualisation that is established during dialogue frequently inhibits the reuse of the same linguistic choice for a different meaning.

Further changes in referential choice might come from task expertise, with speakers in dialogue switching to more appropriate alternatives as they evaluate the fit of their previous expressions to deal with the communicative demands of the task (Healey & Mills, 2006). But likely the single most important source of change in reference choice is the addressees themselves, as a referential expression is only reused if there is sufficient evidence of understanding of its previous use. Usually, this evidence comes in the form of explicit, full-turn feedback, or backchannels, which can either explicitly signal acceptance or indicate that the conversation is to continue, thereby establishing the assumption that previously used expressions have been understood (Clark & Carlson, 1982). Experimental evidence has shown that interrupting feedback channels has important consequences over linguistic choices in dialogue. Healey and colleagues (Healey, Mills, Eshghi, & Howes, 2018; Mills & Healey, 2006) have shown that artificially making feedback requests more general tends to drive speakers towards the use of more abstract conceptualisations, suggesting change in reference choices is at least partially modulated by the perceived certainty of their acceptance by the addressee.

Similarly, the effect of feedback on speakers' reference production has been shown to influence the change and evolution of these references. Different studies have suggested the absence of feedback in the production of references from individuals, compared to pairs of interacting speakers, might explain their maintenance or lengthening of referential phrases, in contrast to the reduction in length that is common in interactive reference (Krauss & Weinheimer, 1966; Hupet & Chantraine, 1992). When referring in natural dialogue, feedback from a partner might amount to the co-construction of a reference, or to its change to a different conceptualisation, as the act of referring is in itself a collaborative process that requires some form of acceptance by the addressee to be completed (Clark & Wilkes-Gibbs, 1986). Brennan and Clark (1996) showed how interacting speakers collaborate in the construction of a conceptualisation they both agree on even when the roles of speaker and addressee (Director and Matcher, in their experimental design) are pre-defined, with addressees even proposing alternative conceptualisations when the speakers' were not satisfactory. Moreover, Brennan and Clark showed how addressees directly influenced the length and level of detail of the speakers' descriptions by using silences or interruptions to mark their understanding or misunderstanding of a reference.

Taken together, these results seem to confirm that the use of conventional references in pairs of speakers is flexible and opportunistic, with speakers adapting to changing contexts by either expanding their previous references or changing them altogether, considering both their own processing needs and those of their partners (Brennan & Clark, 1996; Van Der Wege, 2009).

## **2.6 Relationship between selection at the dialogue level and language change**

The subjects known as *language evolution* and *language change* can be seen as representing different time-frames for a single problem, namely, the dynamic process by which a community transforms its communication system through use, adapting it to their communicative needs. Due to the dynamic nature of language as a phenomenon, the process by which language change is produced, selected for,

and propagated in a population can be said to be a permanent feature of the system, which is reorganised continuously through use (Haselow, 2018). As with the biological evolution of organisms, it might be difficult to identify the moment in which a variant becomes established, since its establishment itself can be better conceived, in a human life time-scale, as a statistical event that can be measured as a change in proportions of use, rather than an absolute fact with a defined timing. On a broader scale, however, as with species in biological evolution, it is easier to establish the appearance or disappearance of a variant, and therefore to analyse the process of linguistic change as such.

However different these two timescales are —the scale of language use, when a variant is produced, and the scale of language change, where a variant is said to have been selected and incorporated into the language—, their interaction constitutes a continuous process, in the sense that linguistic change emerges from the accumulation of instances of selection of variants in language production (Bybee, 2006; Garrod & Pickering, 2013). Therefore, it makes sense to analyse the dynamics of language change in linguistic interaction as a window into the broad process of language change.

As the use of a variant activates a cumulative process of priming (for both the speaker and the addressee), the repetition of the same variant in a conversation as a convention between the interlocutors and the accumulation of instances of use of that same variant in a population can be seen as a continuous process, by which a signal-meaning pairing that is used conventionally in a conversation is acquired by other members of the population through interaction. Malt and Sloman (2004) have shown how reference choice in conversation has long-lasting effects. In their study, participants who had used one name (out of at least two equally likely alternatives) as a reference in interaction maintained the same name in subsequent interactions with other speakers. They argue that each instance of name use strengthens the association between that name and the object it refers to, and that that strengthening is not limited by the end of the dialogue but might extend over multiple interactions in (at least) a span of days (see also Markman & Makin, 1998). Moreover, each of these naming instances contributes to the overall naming preference for that object, as they not only increase the frequency count that the speaker maintains for that reference (a form of statistical learning), but also affect

the cognitive representation of the reference, as it becomes associated to its successful use.

A similar additive effect was observed by Brennan and Clark (1996), who showed that the use of referential expressions was tied to the frequency of their past use, as expressions who had been used more often (in 4 trials vs 1) were maintained significantly more and for longer periods of time in subsequent interaction. This effect was independent of which speaker (Director or Matcher) had originally proposed the conceptualisation. More strikingly, Gurevich, Johnson, and Goldberg (2010) found that speakers remembered and were able to recall sentences after only one incidental exposure (while reading 300-word texts, and with no explicit instruction to memorise), adding to the argument that each instance of use affects representation and subsequent use. In their experiment, participants were more likely to describe a scene using the same sentences that they were exposed to (on a verbatim criterion), compared to alternative sentences with the exact same meaning, even when the testing occurred 6 days after the initial exposure.

Frequency has been also defined by Croft (2013; see also Bybee, 2006) as one of the main drivers in the shaping of speakers' use of grammar. In this sense, a speaker's use of language is at least partly determined by the frequency of words and structures in the language she has been exposed to (including her own production). In this sense, grammatical change can then be seen as the process by which changes in the accumulated frequency speakers are exposed to (due to the accumulation of changes that are naturally produced in language use) change the speakers' own production, which in turn contributes to changing the overall production frequencies for that language<sup>2</sup>. However, as Haselow (2018) suggests, language is not processed as unstructured input, but as concrete communicative experiences, therefore co-occurring regularities and transitional probabilities are incorporated as intrinsic properties which define the grammar an individual uses not

---

<sup>2</sup> Even though actual change in reference choice –the selection of a variant that is not the one your interlocutor produced– is most likely a conscious decision (at least as a reflective process), most instances of language change occur below the consciousness threshold. Different studies have addressed the accumulation of minor changes in pronunciation or phonetic realisation of words and sounds, concluding that frequency is a main driver of change also in this domain. Bybee (2006) has showed how the higher frequency of a word will impact its phonetic realisation by leading to the shortening or attenuation of the articulatory gestures that form that word, thereby affecting both the way the word is uttered and the accumulated frequency of that sound in each speaker's frequency count.



as random frequencies, but as a structured set that is continuously re-structured as we use it.

To the speaker, linguistic structures look relatively stable, as the community appears as a conservative buffer around the most frequently used alternative, shielding the user from the influences of individual changes through the appeal to an established convention, but in the real-time accumulation of minute changes to the system, change is continuously in-the-making (Haselow, 2018). This accumulation builds up over time through routinisation, either in local communities (cultural groups, jargons, trends, etc.) or gradually throughout the language community, where each new alternative that increases its usage might potentially overthrow the previously 'ruling' convention.

However, conventionalisation is not a direct, mechanistic outcome of a high frequency of use. Speakers use language to communicate meanings, and those meanings (incorporating, among other features, the intentionality of the utterance, the identity of the interlocutors, etc., see Brown-Schmidt et al., 2015) are integrated into the dynamic re-organisation of linguistic structure that derives from their accumulation. As conventions are created as solutions to frequently encountered communicative tasks, what is accumulated is not an abstract count but a meaningful measure of using that token to solve a recurrent communication problem.

Experimentally, the relationship between the two timescales –that of language use in interaction and that of language change– can be tackled by comparison between iterated learning experiments and dialogical interaction experiments. While iterated learning recreates the effects of several generations of language users on linguistic structure, interactive experiments show how these changes can be observed in the course of repeated interaction between the same or different speakers (Winters, Kirby, & Smith, 2018). The work of Smith and Wonnacott (2010) and Smith, Fehér and Ritt (2014) can be used to compare these two settings, as they used the same experimental task in iterated learning transmission chains and interacting pairs, respectively. In both experiments, unpredictable variation was eliminated as a consequence of use, with similar changes being brought by repeated interaction within the same pair of speakers and by *vertical transmission* along a series of individual users, where the output of each speaker is used as input for the next

speaker to learn from. Even though the two experiments are not directly comparable, their results suggest the effect of interaction amplifies what vertical transmission would achieve in a much longer time<sup>3</sup>.

Similarly, Tamariz and colleagues have compared iterated chains and dialogue in two conditions of the same experiment (Tamariz, Roberts, Martínez, & Santiago, 2018). In their setup, both interacting pairs and individual speakers in iterated chains had to use an artificial language to designate either round, cloud-shaped figures, or pointy, star-shaped figures. Only in the interacting pairs did the language become more iconic, resembling the attributes of the figures that were described. As iconicity is thought to increase communicative success (particularly in artificial languages that have no past history of use), this move might be seen as the result of interlocutor negotiation of new conventions that better fit their communicative needs. Considering how frequency shapes linguistic structure, we might assume repetition of these type of tasks should strengthen the association between those particular solutions to the problem in the speakers' memory, turning these 'better' signs into communicative conventions.

When brought to the community-level scenario, this preference for 'better', more efficient signs would be related to the evaluation of the sign 'fitness' as the best tool to communicate a meaning in a given context. Tamariz, Ellison, Barr, and Fay (2014) modelled real micro-societies data and concluded that the final distribution of sign use corresponded to a combination of an egocentric bias (a conservative bias to use the same sign the speaker was already using) and a content bias (an extrinsic bias related to the functionality of the sign in use), which could drive a change from the egocentric sign to a different one. Fay and colleagues (Fay & Ellison, 2013; Fay, Garrod, & Roberts, 2008) similarly found that sign selection in small groups was subject only to guided variation, causing signs to become simpler. Variation in larger groups, however, was subject to both guided variation and a content bias, which, on a considerable sized group acting as a selection pool of competing variants, caused usage to select the form that best suited the communicative needs of the broader population. Considering that in linguistic

---

<sup>3</sup> Differences between transmission and interaction might be observable in other features of language, such as syntactic structure. Kirby, Tamariz, Cornish, and Smith (2015) found that interaction alone was not able to promote increased structure in an artificial language (but see Winters et al., 2018, for a different pattern of results).

interaction the initial signs used to communicate a meaning can be interpreted more easily than in graphical communication, this move towards ‘better’ signs can be linked to change in reference choice, where speakers abandon established choices in their search for better alternatives that ease the effort of the interaction in context. An extension of this idea to the language level is put forward by Lupyan and Dale (2010, 2015; see also Perfors & Navarro, 2014) in their Linguistic Niche hypothesis, which argues that the selective pressures that operate on language in use produce languages that are adapted to their circumstances of use, such that “*a particular grammar can be viewed as an adaptation to a particular environment*” (Lupyan & Dale, 2015: 292).

## **2.7 Conventionalisation in non-linguistic communication**

While psycholinguistics has looked at the conventionalisation and change of references in dialogue, another relevant line of research is the conventionalisation and change of non-linguistic signs explored by experimental semiotic studies (see, for example Galantucci, 2005; Garrod et al. 2007). In these studies, interaction has been linked to the evolution and abstraction of graphical signs, such that signs employed later in the interaction would be more abstract/symbolic and show less iconic resemblance to their referents than signs used earlier on. These evolved signs would be simpler in form, but less transparent to outsiders (Garrod et al., 2007; Schober & Clark, 1989). Importantly, the start of this evolutionary process was not necessarily the first sign used to communicate a given meaning, but rather the one which both participants in a pair had converged upon, showing that the process is more closely related to a referential reduction process, operating onto a grounded signal-meaning pair, than to mechanistic repetition of any arbitrary sign.

Galantucci (2005) showed how, when the use of conventional signs of any kind is not permitted, pairs of interacting participants would consider the successful use of any signal-meaning pairing as a precedent for communication in the next games, thereby creating a pairwise convention out of an otherwise arbitrary signal. Galantucci terms this *learning by using*, and links it with the emergence of linguistic conventions in that the successful signals act as precedents for future instances where a similar coordination need is faced. However, in his work, the non-changing

context and the low number of elements that needed to be remembered provided no incentive for evolution or change in the convention (which was generally arbitrary in origin, and therefore not necessarily ideal for the task). Impeded from communicating in any other way, participants will converge on anything that happens to be successful, and will not change it nor perfect it as in Garrod and colleagues' studies. A follow-up study (Galantucci, Rhodes, & Kroos, 2006) revealed that, faced with a changing context, participants' communication systems adapted to become more combinatorial, with participants combining elements from different signs in order to improve their communication success.

In a recent study, Misyak and colleagues (Misyak, Noguchi, & Chater, 2016) showed that, in a communicative context, a convention can be created instantaneously as speakers consider the communicative context when interpreting a signal, and will extract from the same 'sign' a different or opposite meaning, if the context points to a different interpretation. In this sense, it could be argued that the convention is not the sign-meaning pairing, but the conceptualisation that speakers are attributing to the pairing, which could include more than one traditional meaning. The authors refer to this process as 'joint inference', where speakers would jointly interpret the meaning of a communicative signal considering the context, the common ground between them, and the signal space their interlocutor had access to (which would render a different result than interpreting the signal from an individual point of view, even taking precedence into account).

Trying to explain the differential effects of alignment and partner feedback in the establishment of graphical conventions, Fay and colleagues (Fay, Walker, Swoboda, & Garrod, 2018) tackled both factors separately in their experimental design, allowing for an independent evaluation of the contribution of each factor to communication success. Their design used a Pictionary task where participants took turns communicating a meaning from a given set to their partner by drawing on a whiteboard. Participants were either forbidden to use the same signs as their partner (thereby preventing sign alignment) or forbidden to give or receive feedback. Their results showed that sign alignment was more closely related to communication success than feedback, while contingent feedback was key in the abstraction process (sign symbolization). Even though this separation does not occur in natural dialogue (nor are speakers denied the possibility of aligning with their partners), their

results shed a light on the dynamics of conservation and change in the use of referential conventions.

Fay et al.'s (2018) results might be interpreted as favouring a more dynamic and flexible approach to alignment, that adds speakers' evaluation of the context and requirements of the communicative situation to the traditional mechanistic account. Considering alignment is proposed to stem from low-level processes as priming, and the fact that it cannot be inhibited in natural communication as it was in this experimental setting, we could conclude that a drive for conservatism of linguistic choices is central to our communicative success, while adaptation and evolution of forms are a consequence of the back-and-forth processes of dialogical interaction. In this sense, conservatism would be the default, 'safe' choice, while adaptation and change would occur if there is a significant amount of dialogical exchanges between speakers.

In the following chapters, I will describe a set of experiments that were designed to shed light on the interaction between conservatism and adaptation in reference choice. These experiments tackle both the influences of the context over the referential choices of speakers, and the impact that different participant roles, and the history of interaction between them, can have over the prevalence of conventional choices or adaptation in reference production.

## Chapter 3. Interaction promotes the adaptation of referential conventions to the communicative context<sup>1</sup>

### 3.1 Introduction

For people to communicate effectively, they must coordinate their use of language so that speakers express their meanings in ways that their addressees can easily understand. Imagine describing a rendezvous point in a city for an interlocutor. If meeting in central New York, two obvious possibilities suggest themselves –you could exploit the salient grid-structure of Manhattan to describe the meeting place (*“the corner of 55<sup>th</sup> and 7<sup>th</sup>”*) or alternatively use salient landmarks as reference points (*“two blocks south of Carnegie Hall”*). Some of these expressions may be easier than others to produce and understand, as they refer to more salient aspects of the city –for instance, describing a location in Edinburgh’s historic Old Town in terms of grid locations or even city blocks is unlikely to be a successful strategy, since the arrangement of the Old Town streets reflects historical happenstance rather than a grid-like organization.

How do speakers choose among alternatives? Considerable previous research has emphasized the role of precedents and conventions: Speakers tend to re-use choices –in our example, a grid-based versus landmark-based description strategy– that have previously been successful, and with repeated re-use these choices become established as conventions. But this research does not tell us *why* a particular choice was made in the first place. More critically, this emphasis on conservatism in dialogue implicitly assumes a conversational context in which relevant features (e.g., the physical context) do not change, and where no new pressures are imposed over the

---

<sup>1</sup> This chapter has been submitted as an article to *Cognitive Science*, with the authors Lucia Castillo, Kenny Smith, and Holly Branigan. Lucia Castillo designed the study, ran the experiments, and performed the analyses; while Kenny Smith and Holly Branigan acted as supervisors throughout the process and edited the manuscript. The appendix to this chapter was not included in the article submitted.

interaction. In the real world, however, conversational contexts and speaker goals change, so that sometimes a previously successful communicative choice may no longer be optimal, undermining rather than supporting coordination. Instead, the need to coordinate may push speakers towards change –and these new choices may themselves then act as precedents in subsequent interaction.

Successful communication therefore requires a balance between repetition and change, the old and the new. But little is known about the mechanisms that determine this delicate balancing act. Moreover, the importance of interaction for this process is also understudied. Faced with the pressure of a dynamic context, will an individual speaker adapt in the same way as a speaker who is interacting directly with a partner? In this study, we explore how the context of communication shapes referential conventions as they initially emerge, and whether and how these initial conventions change through repetition and/or interaction.

### **3.1.1 Alignment and the development of conventions**

Many studies have suggested that during interactive dialogue, people solve the problem of how to coordinate by re-using their partner's previous linguistic choices. For example, in classic experimental work on coordination in dialogue, experimental participants were required to repeatedly describe locations for each other while navigating two-dimensional grid-like 'mazes' (Garrod & Anderson, 1987; Garrod & Doherty, 1994). When faced with the choice of how to describe her position in the maze, a speaker might choose to refer to a section in the maze as "*the right indicator*" or as "*the one sticking out*", depending on which expression her partner previously used to describe this element. More generally, she might tend to use the same underlying scheme to describe different positions in the same way, for example choosing to refer to positions in terms of column-row coordinates rather than as salient points in a holistic configuration. In such cases, she would align with her partner's precedent with respect to a particular description scheme rather than a particular expression (e.g., using "C4" following her partner's use of "D2"), forming a shared understanding of the situation or *situation model*.

Re-using a partner's previous choice makes sense as a strategy for effective communication. The shared experience of the partner's original successful use of that label or scheme is part of the pair's store of shared knowledge or *common ground* (Stalnacker, 1978; Clark, 1996). Speakers can appeal to this knowledge when subsequently formulating references for their addressee (i.e., engaging in *audience design*; Clark & Carlson, 1981), as it represents a safe option (because both speakers agree on its meaning) and it is easily accessible (because it has been recently activated in the conversation; Brennan & Clark, 1996; E. V. Clark, 1997). With repeated re-use, this reference becomes a local convention between the pair, whereby both partners hold a mutual expectation of its subsequent use (Lewis, 1969/2002). This concordant mutual expectation is also strengthened by higher-order expectations between the speakers: Each speaker knows that the other speaker expects them to use this convention. The alignment of referential choices between interlocutors may also have a basis in an automatic and resource-free priming mechanism that leads to interlocutors converging on shared alternatives at different linguistic levels (lexical, syntactic, semantic, and situation model; Garrod & Anderson, 1987; Pickering & Garrod, 2004), which in turn entails a common interpretation of the task and facilitates mapping the reused expressions to their referents in the world (Shintel & Keysar, 2009).

While these accounts explain how conventions are initially established through interaction, they do not address the two questions which are central to this paper: what determines the form of the initial convention, and when and why are conventions sometimes overturned?

### **3.1.2 Determiners of initial conventions**

Alignment might explain how interacting partners develop a conventionalized scheme for referring to things on the basis of one speaker's initial choice and its subsequent repetition. But it does not address the question of why the speaker initially chose that scheme rather than another.

In the absence of previous linguistic context, the initial selection of a description scheme might plausibly be shaped by the non-linguistic context. For example, in



spatial tasks (such as describing locations in a city or an experimental maze), the perceptual context of the communicative situation might play a role in determining which linguistic alternative is preferred. Specifically, salient landmarks or locations can act as either starting points or references to locate other objects or trajectories, constituting a kind of pre-linguistic common ground between interlocutors (Wilkes-Gibbs & Clark, 1992). If our speaker assumes that her interlocutor shares her evaluation of the salience of specific elements in the maze, she might choose to rely on those concrete landmarks to describe her positions (e.g. *“the long column on the left”, “the one sticking out”*). Alternatively, if there are no such landmarks that she can rely on, or if the environment provides another salient conceptualization (e.g., of the maze as a grid), she might opt for a more abstract approach (e.g. *“first column, third square down”*), depending on her evaluation of the *fitness landscape* that the perceptual context provides for her descriptions. Importantly, this evaluation would yield similar results whether the speaker was interacting directly with a listener or addressing a future (non-present) interlocutor.

Early in an interaction, we should therefore expect environmental factors (perceptual context, speaker identity, etc.) to outweigh historical factors (current or previous interactions between the same pair) in determining description scheme choices. However, as the interaction progresses, referential expressions that have been successfully used as a coordination tool become salient in the eyes of the interlocutors, shifting speakers' evaluation of the task from a pre-linguistic analysis of the context to an analysis of their common history regarding that context (Garrod et al., 2007; Vogels et al., 2013). Successful references then become precedents that speakers can call upon, leading to the development of local conventions. In the same way, an individual speaker lacking concurrent feedback from an audience would become entrained with their own previous expressions, as the self-monitoring process generates priming effects on the speakers' own production that are equivalent to the effects of priming from an interlocutor's speech (Pickering & Garrod, 2004).

If the communicative situation does not change, the conventions that interlocutors develop should therefore be relatively well-adapted to effective communication, because the original description which served as the basis for the convention will generally fit that situation. These initial conventions can then provide a basis for subsequent fine-tuning, which refine or economize initial conventions over the course

of repeated use. A well-known example shows how the descriptions of tangram figures, repeatedly exchanged by interlocutors over the course of a game, maintain their core while dropping other attributes, for example going from “*a man who looks like he might be pushing something to the right*” to “*pushing man*” (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986). Research in graphical communication has similarly shown that the signs developed in interaction are usually abstracted and refined until an optimal trade-off between ease and economy of production and comprehension has been reached (Fay, Ellison, & Garrod, 2014; Fay et al., 2008; Garrod et al., 2007), and even when interlocutors maintain the same conceptual framework, more efficient ways of interaction can be achieved (Garrod & Doherty, 1994; Mills, 2014).

### **3.1.3 Alignment and change**

The accounts of convention-formation reviewed above imply that, once established, interlocutors will not abandon existing precedents in favour of new alternatives: While the form of the conventions might be streamlined, the conceptual pacts they reflect will be preserved. For example, if two experimental participants come to refer to a salient point in a maze as “*the right indicator*”, they will both expect their interlocutor to continue using this convention (Brennan & Clark, 1996; Clark, 1997), assuming the communicative situation remains constant.

But communicative situations frequently do change: New referents may come into play, the perceptual context may alter, or the task itself may impose new demands. As the conversational situation changes, established conventions may no longer be appropriate, and successful coordination might require speakers to adapt their referential behaviour dynamically to better fit the new situation. Speakers can and quite frequently change the way that they refer to referents in natural conversation (Healey, 2008). Nevertheless, such changes come at a cost: Even though previous research on conventions suggests that a pairwise convention should not be as strong as a community convention, and so might be defeasible (as the lack of other individuals privy to the convention eliminates an external pressure for conformity; Garrod & Doherty, 1994; Lewis, 1969/2002), all experimental evidence points to a cost associated with abandoning a precedent (Brennan & Clark, 1996; Kronmüller &

Barr, 2015; Metzling & Brennan, 2003). In such situations, then, the pressure for change conflicts with a pressure to maintain established precedents.

However, experimental evidence has shown that interlocutors are able to dynamically adapt to changing circumstances, taking into account both the context and their shared history of interaction. For example, interlocutors use more informative, disambiguating terms as more similar (and therefore confusable) referents are added to the context (Brennan & Clark, 1996; Van Der Wege, 2009), and can even change the meaning of a conventionalized sign altogether if the information that might be extracted from the context allows for different interpretations of the same sign (Misyak et al., 2016).

Perhaps surprisingly, individuals faced with similar tasks do not seem as flexible as interacting pairs. In both classic figure description studies (Krauss & Weinheimer, 1966) and in graphical communication games (Garrod et al., 2007), individual participants did not reduce the length or complexity of their descriptions over time. Similarly, Van Der Wege (2009) found that speakers producing references to imagined addressees were more likely to overspecify their descriptions than speakers producing references for co-present addressees, and also less likely to adapt their descriptions to the communicative situation. How then would an isolated individual, producing references for a future (imagined) addressee, differ from interacting pairs in dealing with the pressures for maintaining or breaking a precedent? One possibility is that the absence of a co-present interlocutor would relieve the pressure for coordination and therefore free individual speakers of the commitment to a precedent, allowing them to switch easily to better alternatives on a moment-to-moment basis. If this were the case, then individual speakers should adapt more rapidly than interacting pairs to pressures from the context. On the other hand, if adaptability is related to a dynamical evaluation of the fit between the favoured reference and the task, then the feedback that pairs provide to each other should play a more substantial role, helping interacting pairs to adapt more easily than individual speakers.

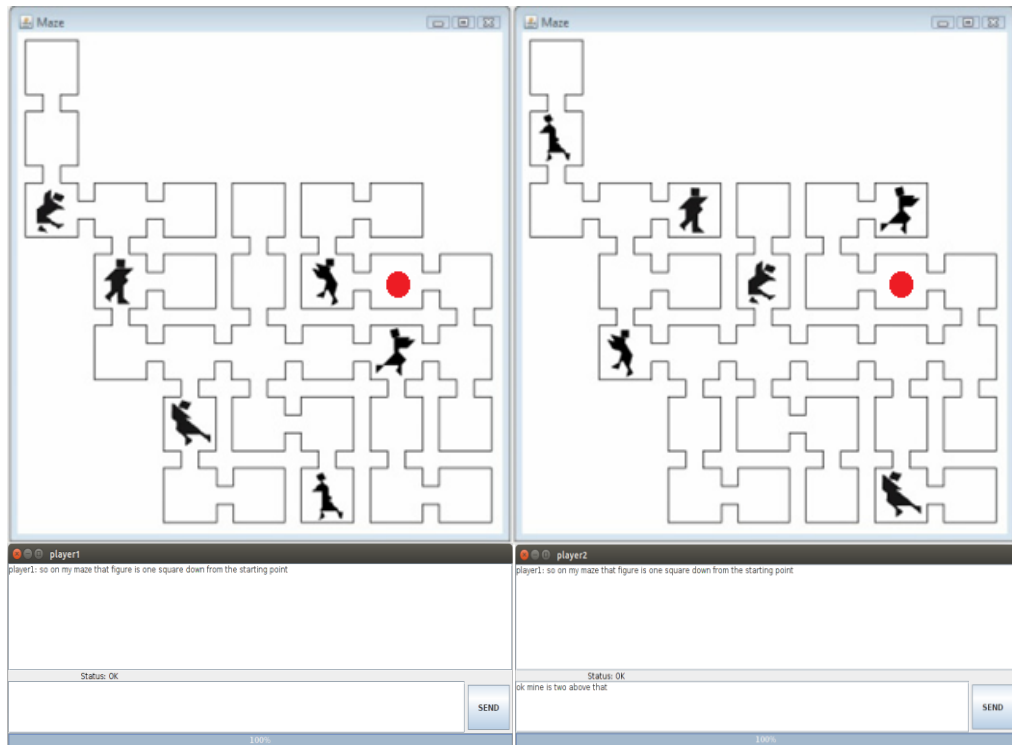
### 3.1.4 The current study

We have argued above that successful collaborative interaction entails a dynamic evaluation that links the speakers' linguistic choices to the demands of the joint action being performed, and to the choices of their co-speaker (Fusaroli & Tylén, 2012). Importantly, these factors may sometimes exert conflicting pressures: A perceptually salient choice might not be communicatively efficient, and repeating a partner's linguistic choice does not ensure optimal adaptation to the task. Thus, the emergence, establishment, and evolution of linguistic conventions are likely to be the product of multiple competing pressures.

We now report an experimental study that addressed these issues. Specifically, we sought to test the influence of non-linguistic context on speakers' linguistic choices, and the effect of interaction over these choices. Additionally, we sought to explore the relationship between communicative alignment<sup>2</sup> and adaptation to the task. We used a maze game (Garrod & Anderson, 1987; Garrod & Doherty, 1994) to provide a simple and controlled context, in which Individuals and Pairs of participants had to solve a recurrent coordination task. The task required participants to communicate the positions of tangram figures, distributed quasi-randomly in a 'maze' (a series of connected squares), to their partners (co-present in the Pairs condition, and an imagined future participant in the Individuals condition) and –in the Pairs condition– to locate on their own mazes the positions that were given by their partners (see Figure 3.1).

---

<sup>2</sup> By 'alignment' we mean here the semantic convergence that participants in the maze game can achieve through the reuse of the same description scheme as their partner previously used. As nearly all possible descriptions in the maze game belong to one out of four different description schemes, alignment will be coded as the reuse (or not) of the same description scheme the partner used in the preceding turn.



**Fig. 3.1** Mazes for Player A (upper left) and B (upper right) in a round of the game, as it appeared to players – players also saw a window featuring the chat tool interface which they could use to communicate with their partner (lower left and right). The player's own position in the maze is indicated by the dot, and there are 6 tangram figures randomly positioned within the maze – note that the two players have the same tangrams, but in different positions. This maze has a high regularity score of 0.821; see Figure 3.2 for irregular mazes.

Participants communicated via an online chat tool, and each participant completed 3 rounds of the game. The maze layout and the tangram figures were held constant across all 3 rounds (thereby repeating the same cues in terms of context and items to be described), but the positions of the tangrams were different at each round (thus forcing participants to describe a changing set of positions in the maze at each round). We manipulated the regularity of the maze layout between participants: Half of the participants played on mazes with a regular grid-like configuration, and the other half played on mazes with an irregular configuration featuring salient sub-components.

The experimental setup therefore presented participants with a recurrent coordination problem: Even though the mazes were structurally the same from round to round, the tangrams' changing positions meant that participants had to describe different positions on each turn. Hence although they could re-use tangram references and an

overall conceptualization of the maze, they could not re-use specific position descriptions. By presenting participants with this recurrent problem, we aimed to push them towards establishing a consistent description system in order to facilitate processing position information, though participants were not explicitly told to do so. Participants in the Individual condition had to sequentially describe tangram positions for a future participant, and hence received no feedback or input in this process; in contrast, participants in the Pairs condition had to take turns describing and confirming with each other the position of each figure in their respective mazes.

Additionally, our manipulation of maze regularity provided spatial contexts that we expected to promote different solutions to this problem. Regular mazes were especially compatible with conceptualizations that emphasized the use of an abstract and invariant element in interpreting positions, namely an imaginary 7x7 grid (Anderson & Garrod, 1987; Healey, 1997). Schemes based on these abstract conceptualizations should be easily generalized to new situations (e.g., new positions within the same maze, or indeed new mazes). In contrast, we expected Irregular mazes to promote the use of conceptualizations that emphasized specific features of the maze or trajectories between salient positions, thus generating schemas that are dependent on the particular disposition of the maze being described and therefore more difficult to generalize (Healey, 1997).

With this set-up, we sought to understand, first, what determines people's initial linguistic choices, and what role does perceptual context play? If people's initial references are influenced by properties of the physical context, then we might expect that Regular mazes, with their visual cues to a consistent underlying structure, would prompt participants to use abstract, systematic descriptions, whereas Irregular mazes, with their visual cues to salient distinct components, would prompt participants to use concrete, figural descriptions (Anderson & Garrod, 1987; Garrod & Anderson, 1987; Garrod & Doherty, 1994). If layout regularity does influence the initial use of Abstract versus Concrete descriptions, this would show that non-linguistic (in this case, spatial/perceptual) context plays an important role in determining how people initially conceptualize, and hence decide how to refer to, the world.

Second, how does interaction affect participants' adaptation to context? If a repeated need for coordination promotes the use of more efficient descriptions, then we would

expect that interacting Pairs of speakers would increasingly come to use descriptions that could be generalized across situations, in other words, they would show an increasing tendency to use Abstract descriptions, even though this might mean abandoning an established precedent; and that they would be faster and require fewer turns to complete the task if they did so. This move towards Abstract descriptions should be particularly marked in speakers playing on Irregular mazes, who might initially use Concrete descriptions (influenced by the maze layout) but then –under the pressure to repeatedly coordinate– should transition to using Abstract descriptions. In contrast, Individual participants working alone, who repeatedly described positions but did not have to coordinate with a partner, should be less likely to change their description strategy. This conservative behaviour of Individual speakers would be expected both under a priming-based model of referential choices, as self-monitoring implies speakers prime themselves through monitoring of their own speech, and under an audience design model, as maintaining the same description scheme would be a good strategy to ensure comprehension by a future interlocutor in the absence of in-the-moment feedback (Pickering & Garrod, 2004; Clark & Carlson, 1981; Horton & Gerrig, 2002).

Finally, and more speculatively, how does linguistic alignment interact with other aspects of communication? Specifically, we explore whether alignment in one aspect of a conversation (exchanging descriptions of positions) would be affected by convergence in a different aspect (jointly constructing a figure's descriptive reference). Thus we tested whether successful referential contraction, where speakers maintain the core of their description but simplify additional details, would be associated with higher alignment levels in the description schemes of Pairs. Including repeated reference to tangrams also acted as a check of our interaction manipulation, by allowing us to compare the canonical tangram effect in Pairs versus Individual participants. Following previous research (Branigan et al., 2011), we anticipated that Pairs of speakers would shorten references to tangrams over consecutive rounds to a greater extent than Individuals.

## **3.2 Method**

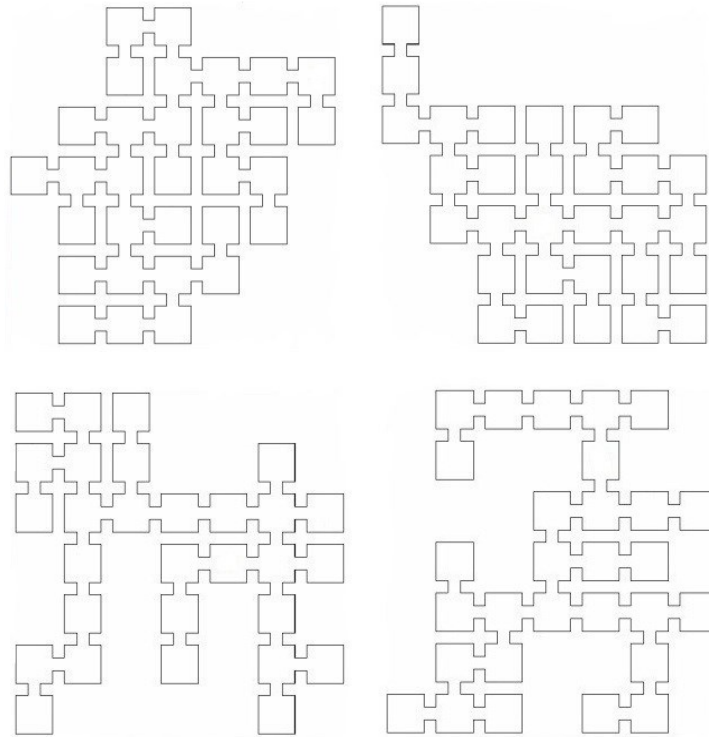
### **3.2.1 Participants**

128 participants, comprising 32 Pairs of participants (52 females, 12 males; mean age 21.5 years) and 64 Individuals (50 females, 14 males; mean age 22.3 years), all University of Edinburgh students, took part in the experiment for payment. All participants were native English speakers. Participants read and signed an informed consent form before taking part in the experiment. 16 Pairs and 32 Individuals were assigned to each maze Layout (Regular vs Irregular).

### **3.2.2 Materials**

We generated four mazes with a maximum size of 7 vertical squares and 7 horizontal squares each (Fig. 3.2). Two mazes were Regular (high regularity score) and two mazes were Irregular (low regularity score). Maze regularity was measured using the following algorithm: For each square in a maze we calculated the proportion of occupied neighbouring squares, considering all 8 (6 for edge squares) surrounding squares (i.e. a square surrounded by occupied squares would obtain a score of 1, a square with only a single neighbour would obtain a score of 0.125), and then took the mean over all squares in the maze to produce a maze regularity score. With this method, mazes that are highly clustered and leave few spaces between squares obtain high regularity scores, while mazes where squares are distributed more loosely throughout the 7x7 space, leaving empty squares in between, obtain lower regularity scores. To ensure our manipulation was linked to regularity differences and not to other properties of the mazes (such as filled vs empty space), Regular and Irregular mazes had the same number of squares ( $\pm 1$ ). Controlling the number of squares in the maze in this way ensures that any differences in description schemes between Regular versus Irregular mazes cannot be attributed to difficulty/cognitive load (cf. Anderson & Garrod, 1987, where a 'regular' maze comprising a full 6x6 grid totalling 36 squares was compared with an 'irregular' maze comprising 24 squares distributed over the same 6x6 grid). Each Individual or Pair was assigned to one of the four mazes, and used this maze in all three rounds.





**Fig. 3.2** Regular (top) and Irregular (bottom) mazes. Regular mazes have regularity scores of 0.864 and 0.821, while Irregular mazes scores are 0.672 and 0.668 (e.g. a regularity score of 0.864 indicates that on average, 86.4% of the neighbours of a square in that maze are occupied, i.e. roughly 7 out of 8). As the figure shows, Irregular mazes contain more spaces between occupied squares, which leads to multiple salient configurational features (protrusions, indentations and clusters).

We also selected 6 tangrams which were to be the target of description in all conditions. These tangrams were chosen from a set of 30 tangrams that was pre-tested to evaluate the dispersion of their emergent descriptions. Using an online survey, we asked 30 participants (all University of Edinburgh Psychology undergraduates, who took part in exchange for course credit) to describe each figure. Tangrams that elicited the use of the same concept in more than 75% of descriptions were discarded, as this implied they would be too easily identified during interaction in the main experiment. The final set (Fig. 3.3) was chosen among the figures that were described as “*dancing*” by at least 20% of participants, which ensured that the tangrams in the set had common features, but that were described using other concepts by at least 50% of participants, allowing them to be adequately differentiated.



**Fig. 3.3** Tangram set

In both conditions participants provided descriptions using a text-based chat tool (based on the DiET chattool software, Healey & Mills, submitted) which was configured to show one turn of dialogue at a time. This restriction was intended to roughly simulate the fading property of spoken dialogue, and to ensure participants would maintain a continuous interaction, instead of delivering all their information in one turn. Each participants' chat window displayed information about the typing status of the other participant (if present), remaining round time, and text of the current dialogue turn. Turns did not fade until the next turn by any participant was submitted.

The experiment code presented the mazes and tangrams and recorded participants' typed utterances, and was programmed using Java.

### **3.2.3 Design**

We used a 2 x 2 x 3 mixed design with the factors Interaction (Individuals vs. Pairs; between-participants); maze Layout (Regular vs. Irregular; between-participants); and Round (Round 1 vs. Round 2 vs. Round 3; within-participants).

### **3.2.4 Procedure**

Participants came to the lab in pairs; we ensured that members of a pair did not know each other in advance. They were randomly assigned to an Interaction condition and maze Layout (Individual vs. Pairs, Regular vs. Irregular maze). They were given verbal and written instructions together, then seated in individual sound-proofed booths equipped with a network computer. In the Individual condition, participants were told that they would be providing written descriptions for a future participant to follow, and that they should aim to make each figure and its position individually identifiable. In the Pairs condition, participants were told they could communicate freely through the online chat tool, that they would exchange descriptions of the

figures and their positions, and that they should aim to make the figure and its position identifiable to their partner.

Individuals and Pairs of participants played three rounds on the same Regular or Irregular maze; they were informed that both members of a Pair would have identical-looking mazes. The mazes contained 6 tangram figures in quasi-random positions; in the Pairs condition, the position of the tangrams differed for each partner, but participants could only see their own figures and placement icon (red dot, see Fig. 3.1), and had no other clues about the position of the figures and icon in their partner's maze. The same tangram figures appeared in each round, but in different positions. All participants had the same starting point in each round, where their placement icon was positioned. Each participant's screen displayed the current maze to be solved, and a chat window showing the last turn of text submitted by any participant and the remaining round time. In the Pairs condition, the chat window additionally displayed a server message that appeared when the other member of the dyad was typing ("*Participant X is typing*").

In each round, participants worked through the set of 6 tangrams in any order they chose. For each tangram, participants had to describe its position so that a future player (Individuals condition) or their current partner (Pairs condition) could move their icon to that position in the maze (by moving through the paths connecting the squares). In the Individual condition, participants chose a tangram and described its shape and position; they then moved to that position and pressed a key, at which point the selected tangram disappeared from the maze, and they then repeated this procedure until they had cleared all tangrams from the maze. In the Pairs condition, participants agreed on which tangram to choose and exchanged descriptions of its location in their respective mazes, such that each player could move to the position where the other player's tangram was located. After placing their icons in these complementary positions, both participants pressed a key to ratify their choice, at which point the selected tangram disappeared from the maze, and they moved on to describe and select a new tangram. If the icon was placed in a wrong location (that is, not where the partner's corresponding tangram would have been located), the tangram figures did not disappear, but no other feedback was given to participants. A round was finished when the participants managed to make all their tangrams

disappear, or when the allocated time of 25 minutes was up (6 Pairs ran out of time in the first round, and 1 Pair in the second round; no Individuals).

### 3.2.5 Coding of transcripts

Participants' utterances were initially coded into three categories: tangram description, position reference, or other (greetings, jokes, etc.). The coding of the description schemes used for position references is based on a simplification of the original four description schemes defined by Garrod and colleagues (Garrod & Anderson, 1987; Garrod & Doherty, 1994): Figurative (references based on salient points or elements, e.g., "*The sticking-up bit on the right*"), Path (references based on a trajectory between a salient point or element and a goal, e.g., "*Two down and one left from the sticking-up bit*"), Line (references based on vertical or horizontal elements in the maze, e.g., "*The second horizontal row, all the way to the right*"), and Coordinate (references established as the intersection of a vertical and a horizontal element, e.g., "*2<sup>nd</sup> column, 4<sup>th</sup> row*"). All the descriptions that participants produced were variants of these four alternatives, sharing their main conceptualizations.

For modelling purposes, each description by a participant was coded as either Concrete or Abstract. Figural and Path descriptions were classified as Concrete ('Type 1 sub-language' in the terminology of Healey, 2008), as they require each position to be described in relation to a specific element in the maze, and are therefore dependent on the actual configuration of the maze. Line and Coordinate descriptions were classified as Abstract ('Type 2 sub-language' in Healey, 2008), as they rely on an abstract grid-like pattern which is independent from the actual configuration of the maze, and could therefore be applied to any maze without change. All position descriptions were coded, totalling 1634 descriptions from 32 Pairs, and 1035 descriptions from 64 Individual participants. Descriptions that were 'incomplete', that is, not containing enough data to single out a specific maze square, were coded considering the conceptualisation they depended upon to obtain full meaning (e.g. "*on the same column*" would be coded as Abstract, while "*the second one after that*" would be coded as Concrete).

Each position description was coded for alignment, i.e. whether it used the same description scheme (Figurative, Path, Line or Coordinate) as the preceding description produced by the other participant, in the Pairs condition, or by the same participant, in the Individual condition (i.e. a Figural description followed by another Figural description would be counted as alignment, but a Figural description followed by a Path description would not); we used this finer-grained set of contrasts to obtain a more detailed appreciation of convergence over time. For Pairs, if the same speaker made more than one consecutive position reference, the comparison was nonetheless performed against the partner's previous reference (i.e., we did not consider self-alignment in Pairs). For Individual participants, alignment was measured as reuse of their own previous description scheme.

Timestamps were recorded automatically by the software. Time per round was measured from the onset of the first typing activity by any participant, to the submission of the last utterance in that round. Rounds were ended when participants registered a correct selection for their final tangram, or when the allocated limit of 25 minutes was reached.

Length of tangram descriptions (total number of words per description) was measured by Pair, since it was unlikely that a participant would describe a tangram that had already been described by their partner in the same round (for the same reason, we could not analyse alignment in tangram descriptions). However, since descriptions were frequently co-constructed by both participants in a Pair (particularly in the first rounds), all descriptions related to the same tangram in a given round were registered as one, no matter if they came from one participant or both. An example of a co-constructed description is given in Box 1.

A: I have one where <b>the character has both hands facing to the right</b> A: <b>and kinda is skipping a bit</b> B: i think i have one like that, <b>looks a bit like it's lunging</b> A: yea that's the one
--

**Box 3.1** Example of co-constructed tangram description by two participants, A and B. All words in bold are counted as one description.

### **3.2.6 Differences between written and oral referential communication**

Text chat is a common medium of communication in most of the industrialised world. Millions of people interact through written chat daily, either through a computer or a smartphone. For university students (our experimental cohort), this is a medium they are highly familiar with, and which they use regularly, especially for informal/social interactions. However, this does not rule out the possibility that there are differences in the way references are built or maintained in written dialogue, compared to a more natural face to face interaction.

Text chat has been described as interactionally incoherent (Herring, 1999), as the interactions between speakers are usually fragmented, agrammatical, and interactionally disjointed. However, our experimental task promotes a type of goal-directed dialogue where participants need to take turns in order to fulfil the task, creating a framework in which the structure of the dialogical exchange is partly determined by the nature of the task. Particularly, the issue of feedback, which might substantially influence speakers' referential choices in oral dialogue, is constrained by the written medium in two ways: On one hand, most written communication formats, including the software used in this experiment, only deliver to the addressee any full turns submitted by the speaker, unlike the real-time flow of speech that characterises oral communication. In this sense, there is no opportunity for the addressee to influence the speaker's choices before these choices are expressed in a full turn. On the other hand, written chat neutralises almost any form of extra-linguistic information that the speakers could employ or interpret as feedback, as even delays in responding (which could be considered as equivalent to silences in oral communication) cannot be clearly interpreted communicatively. Taken together, these issues point to a more controlled exchange that might be less pervious to partner influences than oral dialogue.

Even though these differences might subtly affect the speakers' experience in the exchange, the software we have used in this experiment, the DiET chat-tool, has been used extensively in written versions of known experimental tasks, and all results to date can be considered equivalent to their spoken-version counterparts. According to Healey and Mills (submitted), results from different tasks such as the tangram task (Clark & Wilkes-Gibbs, 1986) and the maze task (Garrod & Doherty, 1994) show that

DiET chat samples have local and global patterns that are comparable to those observed in the original spoken versions.

### **3.3 Results**

We present our results in four sub-sections. We first consider how our main experimental manipulation (Regular vs Irregular maze layouts) influenced participants' use of Abstract vs. Concrete descriptions. We then investigate the relationship between their use of Abstract vs. Concrete descriptions and their time taken to complete the task. We next examine how participants' use of descriptions was influenced by their partner's previous behaviour (i.e., whether they aligned on the same description scheme their partner had used before). Finally, we examine how participants' tangram descriptions were affected by interaction and repetition.

#### **3.3.1 Choice of description schemes**

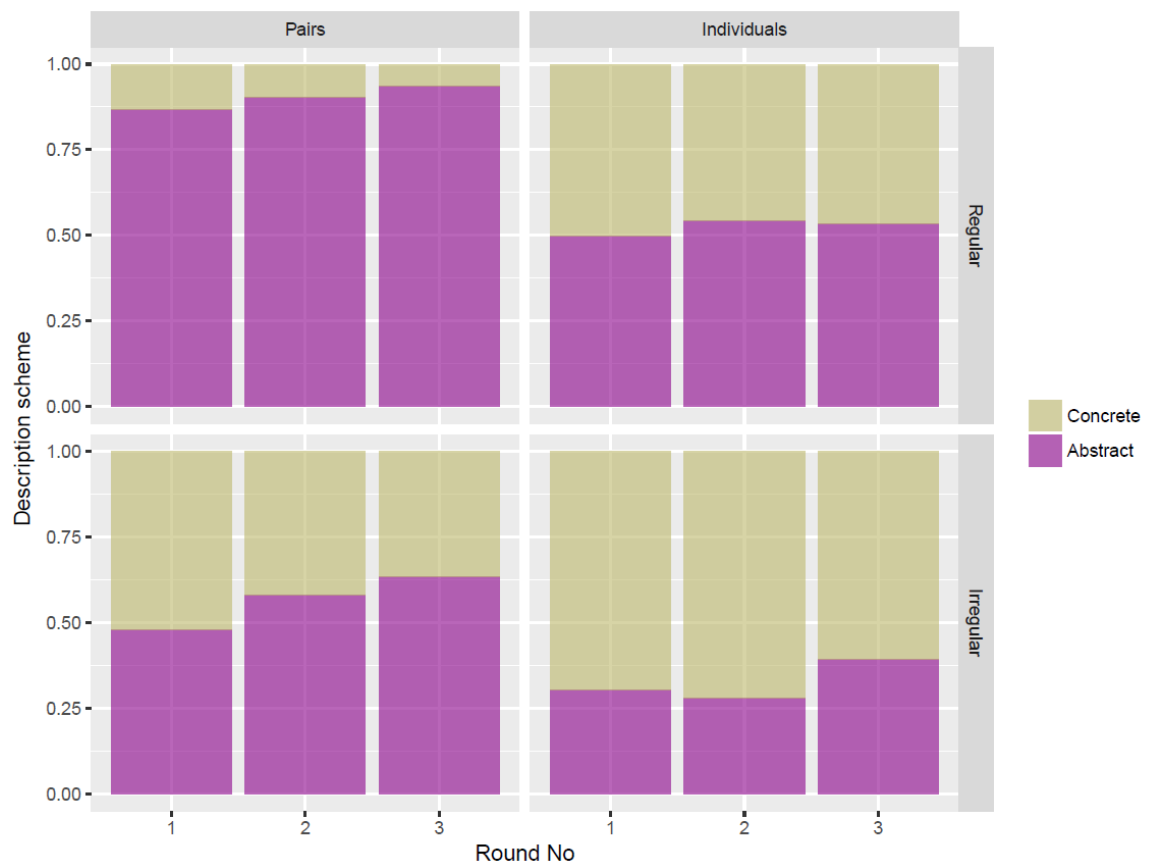
Our experiment presented Pairs and Individual participants with 3 rounds of mazes in either a Regular or an Irregular maze layout. Figure 4 shows the proportion of use of Abstract versus Concrete descriptions in the three rounds of the game for Pairs and Individual participants in Regular and Irregular mazes.

We used a logistic multilevel model (lme4 package in R, Bates, Maechler, Bolker, & Walker, 2015) to evaluate the effect of our layout manipulation on the use of Abstract or Concrete description schemes. We included Interaction condition (Pairs vs Individuals), Layout regularity and Round number as fixed effects, plus their interaction. The random structure was the maximal structure justified by our design (following Barr, Levy, Scheepers, & Tily, 2013), including random intercepts for Pair/Individual and for Participant nested in Pair<sup>3</sup> (to account for the initial variation between pairs and between subjects), and random slopes for Pair/Individual over Round number (to account for the differential effect of round number over the different Pairs or Individual participants). P-values were obtained through likelihood ratio tests

---

<sup>3</sup> Individual participants were assigned a unique 'Pair' number, to allow for the fit of the same model over Pairs and Individual participants.

for each parameter against a model without that parameter, using the function *mixed* in the *afex* package (Singmann, Bolker, Westfall & Aust, 2017).



**Fig. 3.4** Proportions of positional descriptions which used Concrete (light shaded) vs. Abstract (dark shaded) description schemes, over three rounds of interaction, in the Pairs (left panels) and Individuals (right panels) conditions, for Regular mazes (upper panels) and Irregular mazes (lower panels).

There was a significant main effect of Interaction condition (Pairs vs Individuals) on the use of Abstract vs. Concrete descriptions ( $\chi^2=14.00$ ,  $p<0.01$ ): Participants in Pairs were more likely to use Abstract description schemes than Individual participants. Critically, there was a significant effect of Layout ( $\chi^2=10.86$ ,  $p=0.001$ ): Participants tended to use significantly more Abstract descriptions in Regular layouts than Irregular layouts. Round number also had a significant effect on the use of Abstract descriptions ( $\chi^2=7.14$ ,  $p=0.008$ ), with participants using more Abstract descriptions as they played more rounds. Moreover, we found a significant interaction between Interaction condition (Pairs vs Individuals) and Round number ( $\chi^2=8.87$ ,  $p=0.003$ ): Participants in the Pairs condition were significantly more likely to



use Abstract descriptions as they played more rounds, compared to the Individual condition. No other interactions were significant.

**Table 3.1 Results of model of description scheme use**

Predictor	Log-odds Betas	SE	X <sup>2</sup> and p values
Intercept	-0.2575	0.6904	-
Interaction condition	4.7828	1.6181	14.00, p=0.0002***
Layout	4.4647	1.5353	10.86, p=0.0010**
Round number	0.8925	0.2581	7.14, p=0.008**
Interaction condition * Layout	0.4912	2.6813	0.03, p=0.86
Interaction condition * Round number	1.9346	0.8422	8.87, p=0.003**
Layout * Round number	0.7604	0.7237	1.62, p=0.20
Interaction condition * Layout * Round number	0.4779	1.0304	0.21, p=0.65

Log-odds betas, Standard Error, Chi-squared and P-values from Generalized Mixed Model fitted using *mixed* function from the *afex* package. Intercept represents grand mean at Round 1. Model is fitted with deviation-coded variables Interaction condition (Individuals-Pairs) and Layout (Irregular-Regular). Round is coded as a numeric predictor, therefore model shows a linear effect of increasing 1 round.

### 3.3.2 The effect of maze layout and description scheme on time taken to complete the task

Time analyses were conducted separately on Pairs and Individuals, due to the different time demands of the task for the two conditions (exchanging descriptions in Pairs, producing descriptions individually in Individuals).

For the Pairs condition, we ran a linear model to predict the overall time taken to complete the task (i.e. all 3 rounds) from the Layout participants were assigned to and their Use of abstract descriptions (using the mean proportion of Abstract descriptions

produced by each Pair as a numeric value<sup>4</sup>), plus their interaction. To avoid a high correlation between the main predictor and the interaction term, Use of abstract descriptions was centred on its mean. Layout was deviation-coded.

In the Pairs condition, there was no main effect of Layout ( $F(1,28)=2.3217, p=0.138$ ). There was however a significant main effect of Use of abstract descriptions ( $F(1,28)=12.946, p=0.001$ ), with participants who used more Abstract descriptions being faster overall. Moreover, there was a significant interaction between Layout and Use of abstract descriptions ( $F(1,28)=4.3293, p=0.046$ ), suggesting that participants who used more Abstract descriptions in Regular layouts were significantly faster than participants who used more Abstract descriptions in Irregular layouts.

To test the possibility that time taken to complete the task was affected by alignment, we ran a further model that predicted total time per Pair (3 rounds) from Use of abstract descriptions, Alignment rate per Pair, and their interaction. There was no main effect of Alignment rate on the time taken to complete the task,  $F(1,28)= 0.0666, p=0.7982$ .

**Table 3.2 Results of model of time taken to complete task, Pairs condition**

Predictor	Betas	SE	F and p values
Intercept	35.63	104918	-
Layout	5.32	209835	(1,28) 2.3217, $p=0.138$
Use of abstract descriptions	-22.54	375874	(1,28) 12.946, $p=0.001$ ***
Layout * Use of abstract descriptions	-26.06	751748	(1,28) 4.3293, $p=0.046$ *

Betas and Standard Error (from linear model *lm* in the *lme4* package), and Chi-squared and P-values (from likelihood-ratio tests using *anova* function). Time estimates are shown in minutes. Model is fitted with deviation-coded variable Layout (Irregular-Regular). Use of Abstract Descriptions is mean-centred.

In the Individual condition, there were no significant main effects of Layout ( $F(1,51)=2.0428, p=0.159$ ) or Use of abstract descriptions ( $F(1,51)=0.02, p=0.888$ ),

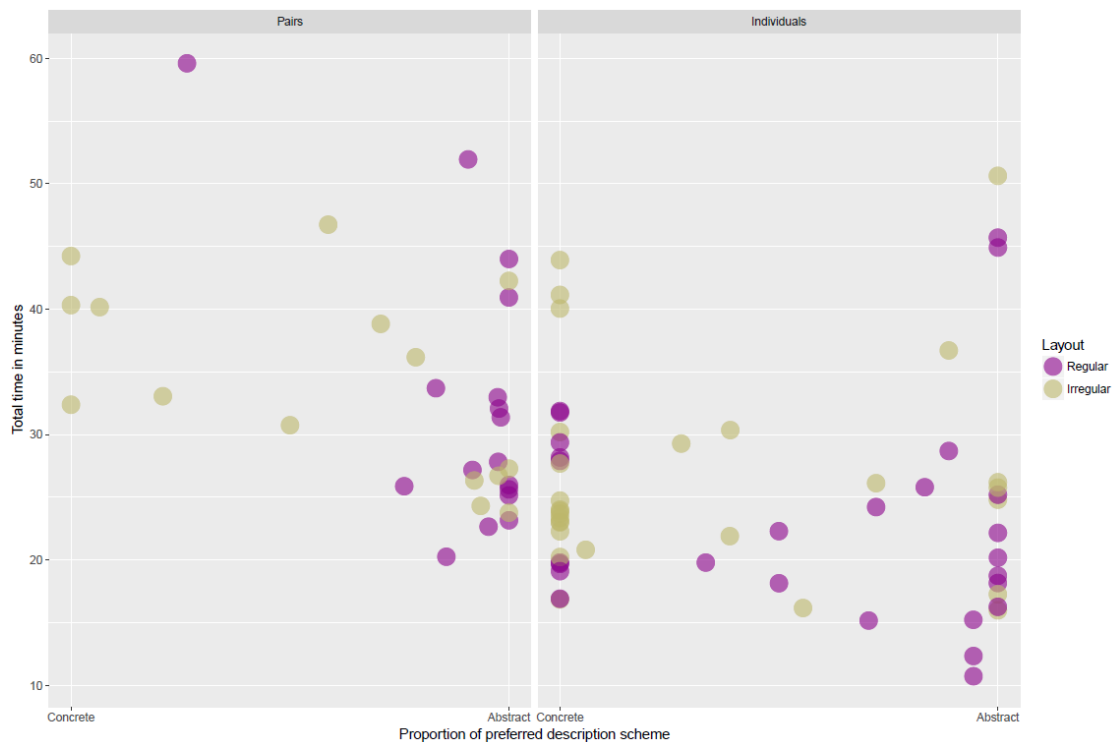
<sup>4</sup> Concrete descriptions were coded as 0 and abstract descriptions as 1. The numeric value used in this model is the mean value of the Abstract column per Pair/Individual, which goes from 0 to 1, reflecting the use of more Concrete descriptions (values closer to 0) or more Abstract descriptions (values closer to 1).

nor was there a significant interaction between Layout and Use of abstract descriptions ( $F(1,51)=0.1147, p=0.736$ ).

**Table 3.3 Results of model of time taken to complete task, Individual condition**

Predictor	Betas	SE	F and p values
Intercept	25.69	73877	-
Layout	-3.51	147754	(1,51) 2.0428, $p=0.159$
Use of abstract descriptions	-0.39	166783	(1,51) 0.02, $p=0.888$
Layout * Use of abstract descriptions	-1.88	333566	(1,51) 0.1147, $p=0.736$

Betas and Standard Error (from linear model *lm* in the *lme4* package), and Chi-squared and P-values (from likelihood-ratio tests using *anova* function). Time estimates are shown in minutes. Model is fitted with deviation-coded variable Layout (Irregular-Regular). Use of abstract descriptions is mean-centred.



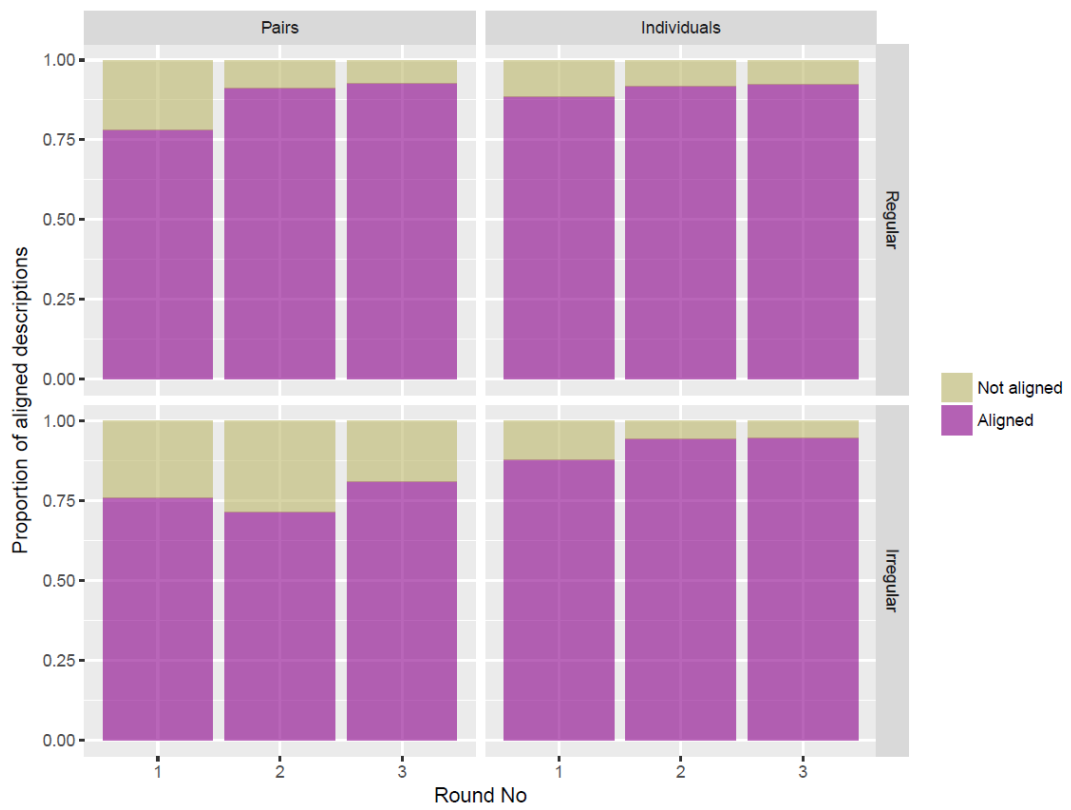
**Fig. 3.5** Time taken to complete the experiment (i.e., 3 rounds) for Pairs (left panel) and Individuals (right panel), as a function of the proportion of Abstract and Concrete descriptions they produced (x axis goes from absolute use of Concrete descriptions on the left, to absolute use of Abstract descriptions on the right, for each panel). Each point represents an Individual or Pair, and points are coloured according to the maze Layout they played on (Regular: dark shaded, Irregular: light shaded).

### 3.3.3 Alignment of description schemes

We ran a binary logistic mixed-effects model predicting alignment on position description schemes (i.e., whether participants used the same description scheme as used in their partner's previous description, in the Pairs condition; or their own previous description, in the Individual condition) based on Interaction condition, Layout, Round number, and their interaction; we used the same random effect structure and coding scheme as reported in section 3.3.1.

We found a main effect of Interaction condition ( $X^2=6.87$ ,  $p=0.009$ ): Individual participants were more highly aligned than participants in Pairs (though recall that Individual participants were self-aligning, whereas Pairs were aligning with their interlocutor); and a main effect of Round number ( $X^2=22.34$ ,  $p<0.001$ ): Alignment scores increased significantly with each round played. No interactions were significant (see table 3.4).

An additional analysis of the data for each Condition separately revealed a differential response by Pairs and Individuals. Whereas Individuals showed only a significant main effect of Round number ( $X^2=15.12$ ,  $p<0.001$ ), Pairs showed both a main effect of Round ( $X^2=9.29$ ,  $p=0.002$ ) and a (marginal) interaction between Layout and Round ( $X^2=3.87$ ,  $p=0.05$ ), with Pairs in Regular layouts showing a greater increase in alignment as they played more rounds, compared to Pairs in Irregular layouts. Although this effect was not strong enough to appear in the full model, it reflects the difference in the alignment trajectory of Pairs according to their initial conventions: Pairs in Regular layouts maintained their initial conventions, consistently increasing their alignment levels as they played more rounds, but Pairs in Irregular layouts abandoned their initial descriptions, temporarily misaligning as a consequence (see Fig. 3.6, Irregular layout).



**Fig. 3.6** Proportions of non-aligned (light shaded) and aligned descriptions (dark shaded) over three rounds of interaction, in the Pairs (left panels) and Individuals (right panels) conditions, for Regular mazes (upper panels) and Irregular mazes (lower panels).

**Table 3.4 Results of model of aligned descriptions**

Predictor	Log-odds Betas	SE	X <sup>2</sup> and p values
Intercept	2.3392	0.2590	-
Interaction condition	-1.1605	0.4568	6.87, p=0.009**
Layout	0.3055	0.4435	0.47, p=0.49
Round number	1.2883	0.3534	22.34, p<0.0001***
Interaction condition * Layout	0.4678	0.8860	0.28, p=0.60
Interaction condition * Round number	-0.3060	0.4211	0.57, p=0.45
Layout * Round number	0.4985	0.3994	1.62, p=0.20
Interaction condition * Layout * Round number	1.2866	0.7883	2.62, p=0.11

Log-odds betas, Standard Error, Chi-squared and P-values from Generalized Mixed Model fitted using *mixed* function from the *afex* package. Intercept represents grand mean at Round 1. Model is fitted with deviation-coded variables Interaction condition (Individuals-Pairs) and Layout (Irregular-Regular). Round is coded as a numeric predictor, therefore model shows a linear effect of increasing 1 round.

### 3.3.4 Tangram description length

Figure 7 shows the average tangram description length per round, measured in number of words per description; within each round we plot the description length of the 1<sup>st</sup> to 5<sup>th</sup> tangram described; the 6<sup>th</sup> tangram was typically not described, since it was by that point the only remaining tangram.

We ran a linear mixed-effects model to predict the (log-transformed) number of words used per tangram description per Pair<sup>5</sup> or Individual participant, with Interaction

---

<sup>5</sup> Because participants either built a joint description of each tangram (more likely in the first rounds than later rounds), or simply agreed with/accepted their partner's description, number of words per tangram is measured per Pair.

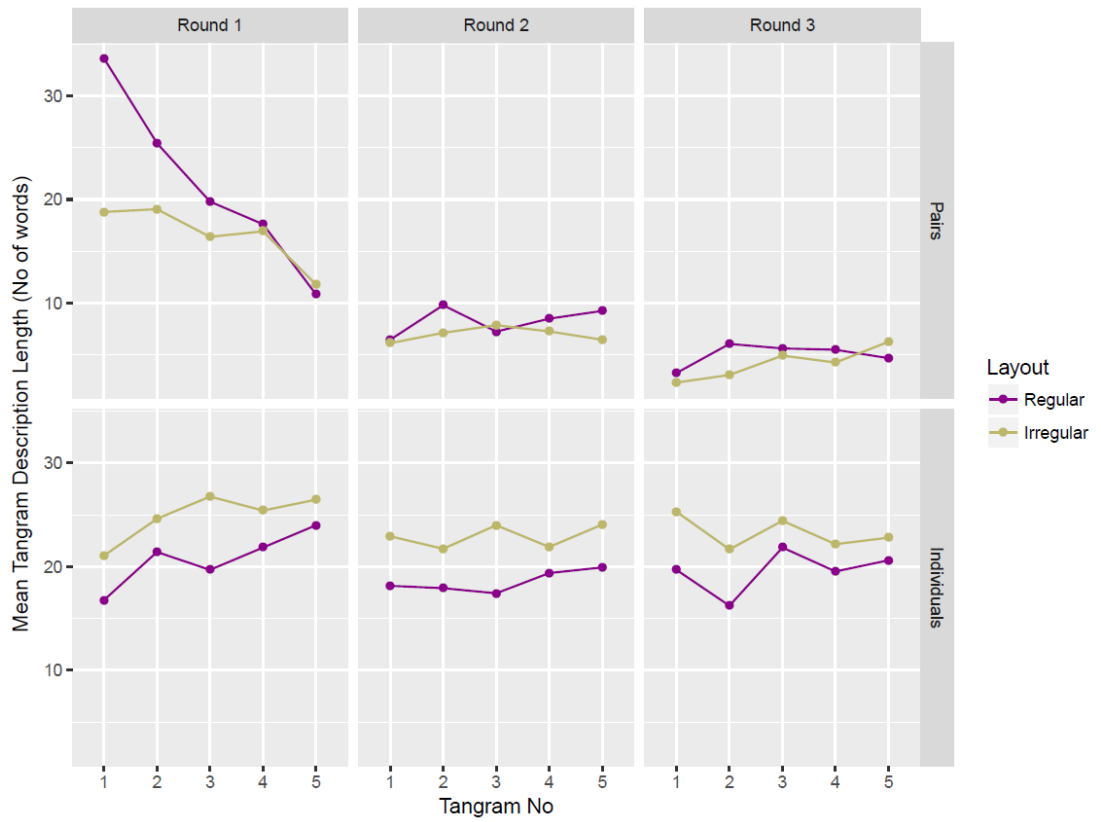
condition, Round number, Layout regularity, Tangram number (1 to 5<sup>6</sup>) and their interactions as fixed effect predictors. We included a random intercept for Pair/Individual and a random slope for Pair/Individual over Round number (the inclusion of a random slope for tangram number resulted in failure to converge). P-values were obtained from the *mixed* function in the *afex* package.

We found a significant main effect of Round number ( $F(1,243)=144.62, p<0.001$ ), with participants using fewer words to describe tangrams with each new round played; and a significant main effect of Tangram number ( $F(1,781)=6.10, p=0.01$ ), with participants using fewer words to describe each additional tangram within each round. There were no main effects of Interaction condition or Layout. Importantly, we found a significant interaction between Interaction condition and Round number ( $F(1,243)=168.82, p<0.001$ ): With each new round played, Pairs' descriptions shortened while Individuals' did not. The interaction between Interaction condition and Tangram number was also significant ( $F(1,781)=25.35, p<0.001$ ): Descriptions of each new tangram in round 1 were shorter in the Pairs condition, but not in the Individual condition. Notice however that there was also a 3-way interaction between Interaction condition, Round number, and Tangram number ( $F(1,778)=31.82, p<0.001$ ): Pairs' descriptions were shorter for each new tangram on round 1, but this effect did not appear in later rounds.

There was also a significant interaction between Round number and Tangram number ( $F(1,778)=13.55, p<0.001$ ): tangram number had a larger effect at round 1 than at later rounds, since descriptions shortened during round 1 but were relatively flat within subsequent rounds; and between Interaction condition and Layout ( $F(1,93)=6.01, p=0.02$ ): Pairs in Regular layouts produced longer descriptions than Pairs in Irregular layouts in round 1, tangram 1, compared to Individuals. As this effect disappeared after a few descriptions, it suggests the additional linguistic coordination needed to establish an abstract description system initially affected participants' verbosity.

---

<sup>6</sup> Participants decided arbitrarily the order of the tangrams they described; there was no pre-established order. Numbering corresponds to the arbitrary order imposed by each Pair in a given round.



**Fig. 3.7** Mean number of words per tangram description per round, in Pairs (upper panels) and Individuals (lower panels), for Regular (dark shaded) and Irregular (light shaded) Layouts.



**Table 3.5 Results of model of tangram description length**

Predictor	Betas	SE	F and p values
Intercept	2.888795	0.0810	-
Interaction condition	0.039335	0.1621	(1,93) 0.06, p=0.81
Layout	-0.067410	0.1621	(1,93) 0.17, p=0.68
Round number	-0.488446	0.0406	(1,243) 144.62, p<0.001***
Tangram number	-0.048489	0.0196	(1,781) 6.10, p=0.01**
Interaction condition * Round number	-1.055468	0.0812	(1,243) 168.82, p<0.001***
Interaction condition * Tangram number	-0.197663	0.0392	(1,781) 25.35, p<0.001***
Round number * Tangram number	0.055621	0.0151	(1,778) 13.55, p<0.001***
Interaction condition * Layout	0.795354	0.3243	(1,93) 6.01, p=0.02*
Layout * Round number	0.016914	0.0812	(1,243) 0.04, p=0.84
Layout * Tangram number	-0.014376	0.0392	(1,781) 0.13, p=0.71
Interaction condition * Layout * Round number	-0.036401	0.1624	(1,243) 0.05, p=0.82
Interaction condition * Layout * Tangram number	-0.076559	0.0784	(1,781) 0.95, p=0.33
Interaction condition * Round number * Tangram number	0.170473	0.0302	(1,778) 31.82, p<0.001***
Layout * Round number * Tangram number	-0.003211	0.0302	(1,778) 0.01, p=0.92
Interaction condition * Layout * Round number * Tangram number	-0.014080	0.0604	(1,778) 0.05, p=0.82

Betas, Standard Error, F-statistics and P-values (from *mixed* function in the *afex* package). Descriptions are considered to be co-constructed by both members of a Pair, therefore each data point is derived from a Pair (in Pairs condition) or Individual (in the Individual condition). Intercept represents grand mean at Round 1 and first tangram. Model is fitted with deviation-coded variables Interaction condition (Individuals-Pairs) and Layout (Irregular-Regular). Round and Tangram number are coded as numeric predictors, therefore model shows the linear effect of increasing 1 round/tangram.

In a final exploratory analysis, we carried out a partial correlation between mean length of tangram descriptions and alignment score (on positional descriptions) per round, controlling for round number, to investigate whether there was a relationship between the referential contraction process and participants' alignment with their partner's previous description scheme use (Pairs condition) or their own previous use (Individual condition). This revealed no significant relationship between mean length of tangram descriptions and alignment score per round ( $r=-0.051$ ,  $p=0.628$ ).

### **3.4 Discussion**

A substantial body of research has considered how dialogue promotes conservatism, with speakers tending to repeat their own and others' previous choices. In contrast, there has been much less consideration of why these choices were made in the first place, why a nascent convention may sometimes be abandoned in favour of an alternative, and more generally how speakers in dialogue balance competing pressures for conservatism and innovation. We addressed these questions through a task in which Pairs and Individuals faced a recurring coordination problem, requiring them to repeatedly describe the position of tangram figures within mazes that differed in their spatial regularity. We examined two aspects of participants' referential behaviour: how they referred to individual tangrams (i.e., choice of an [independent] referential expression), and how they referred to positions in the maze (i.e., choice of a systematic description scheme).

Our results showed that when interacting Pairs of participants repeatedly referred to the same tangrams, they reduced the length of their referential expressions, whereas Individual participants did not. When participants (both interacting Pairs and Individuals) referred to positions in the maze, their choice of description scheme was influenced by the regularity of the maze layouts: Overall, participants used significantly more Abstract descriptions when faced with Regular mazes than when faced with Irregular mazes. However, Pairs and Individuals differed in the consistency with which they used one or other scheme: Whereas Pairs used increasingly more Abstract descriptions as they played more rounds, Individuals maintained their initial choice throughout the 3 rounds of the game.

These differences in description scheme use yielded associated differences in rates of alignment (i.e., the tendency to re-use the same description type across adjacent descriptions), with Individuals showing higher alignment than Pairs. Although both Pairs and Individuals showed an overall increase in alignment as they played more rounds of the game, our data offer some suggestion that Pairs' alignment was mediated by the layout in which they were playing: Pairs in the Regular layouts showed increased alignment as they played more rounds, but Pairs in the Irregular layouts did not (reflecting a switch from an initial preference for Concrete descriptions to a preference for Abstract descriptions).

The regularity of maze layouts did not affect the overall time taken to complete the task. However, Pairs who used more Abstract descriptions were significantly faster to complete the task than Pairs who used more Concrete descriptions; this was particularly the case for Pairs in Regular layouts. Individuals showed no significant differences in time taken to complete the task, for any of our measures. We now consider our results and their theoretical implications in turn.

#### **3.4.1 Referential contraction and the role of interaction**

We start by considering participants' references to tangrams. Our task involved repeated references to the same tangrams in the same way as many previous studies. We found a pattern of effects that is consistent with previous findings. In particular, Pairs showed a canonical referential contraction effect (Barr & Keysar, 2002; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Krauss & Weinheimer, 1964): Descriptions became shorter and simpler but maintained their core (usually a noun phrase) as the interaction unfolded, i.e., an indefinite description such as *"the one where the shape looks like its kinda sliding or skiding"* [sic] in round 1 turned into the standard noun phrase *"the sliding one"* in round 3. In contrast, Individual participants showed no referential contraction and maintained the same references they had been using on round 1, verbatim in most cases, until the end of the game. This disparity between the behaviour of interacting Pairs versus Individual participants replicates previous findings (Branigan et al., 2011) in the context of a more complex task, and confirms the effectiveness of our interactivity manipulation in this context.

Importantly, Pairs' tendency to contract their tangram references was not affected by alignment or misalignment in their position descriptions. A Pair could be transiently misaligned, i.e., in the process of abandoning a concrete description scheme and moving towards an abstract one, while at the same time maintaining their coordinated referential contraction process by converging on the same tangram descriptions (reflected in the absence of a correlation between alignment rate and tangram description length). This pattern reflects an interesting dual processing of different semantic stimuli, where participants can process, contract, and converge on one set of referential expressions (in this case, the descriptions of the tangrams themselves) while at the same time misalign and diverge on a different set of expressions (in this case, the descriptions of the tangrams' positions).

The marked differences between Pairs' and Individuals' behaviour on tangram description can be explained with respect to the absence of feedback in the Individual condition, and the associated provisionality of the Individuals' references. In interacting Pairs, references are proposed by one of the speakers and confirmed (explicitly or implicitly) by their partner as the interaction unfolds, establishing a conceptual pact (Brennan & Clark, 1996). This confirmation determines the certainty of the common ground between the speakers, allowing them to build over that ground by either shortening or simplifying the expressions (Clark & Wilkes-Gibbs, 1986). In this sense, references for an Individual speaker who has no feedback from their assumed partner are always provisional, lacking the basis for shortening or simplifying as the comprehension and acceptance of their expressions is not ratified by an addressee. We note that this lack of confirmatory feedback could also explain the lengthening of individuals' descriptions in the first round, as the basis for reducing descriptions during a round is the simplification of the task by virtue of an agreement between speakers on the figure-description pairing, which is confirmed in the Pairs as both speakers select the same figure. The next description will then be constructed from the pool of remaining tangrams. Even though both Pairs and Individual participants were instructed to physically remove the tangrams from the maze after description, only Pairs seemed to benefit from this as they can be certain their previous description had been understood. Individuals, lacking confirmation of the success of their past descriptions, seemed to put more effort (in terms of number of words) into subsequent descriptions.

### 3.4.2 The influence of context in the emergence of conventional references

Our results provide striking evidence that the perceptual context of interaction can significantly affect the linguistic choices of speakers. Participants who were playing on mazes with differing layouts (Regular or Irregular) initially adopted different description schemes, responding to specific properties of those contexts. In the Regular layout condition, where the mazes were characterized by an ordered, grid-like appearance, participants overwhelmingly chose Abstract descriptions to refer to specific positions; in the Irregular layout condition, in contrast, where the mazes were characterized by irregular protrusions and salient elements, participants were more likely to opt for Concrete descriptions.

Both choices made use of the perceptual salience of the context, indicating an interaction between participants' need for mutual understanding and their individual evaluation of the context's properties. Perceptual salience is generally assumed to emerge from general properties of the object at hand, hence speakers looking at the maze could assume that any other person looking at the same maze would perceive the same elements to be salient, and therefore that those elements should be particularly relevant in communication (Tarenskeen et al., 2015). Perceptual salience therefore acted as a common base or pre-linguistic common ground for the establishment of references. In speakers' initial choice of description schemes, this perceptual common ground modulated the salience of the different scheme choices, by linking the salient properties of the context with a description that reflected that salience.

Each description scheme is furthermore associated with a particular mental model of the maze, which becomes 'agreed upon' when communication is effective (Anderson and Garrod, 1987; Garrod and Anderson, 1987). However, these mental models do not arise directly from participants' visual perception of the maze, but rather from an interpretation of the maze's layout in a way which highlights a specific aspect or property. That is, two players sharing a description scheme are not only sharing a linguistic description but also choosing to focus on a specific property of the maze as the key to the interpretation of both their own descriptions and their partners'.

Our results suggest this adaptation or ‘fit’ between description schemes and maze layouts was taken into account by our participants, as they evaluated the best way to convey their position to a (co-present or future) partner in a specific context. As speakers usually aim for efficient descriptions that would require minimal expansion or repair (considering both the Maxim of Quantity; Grice, 1975, and the influence of perceptual salience; Brennan & Clark, 1996), these different maze layouts might create different fitness landscapes, each favouring specific description schemes over others. However, in contrast to the fixed and unchanging pressures that the concept of ‘landscape’ would suggest, the pressures from the context in our task are not static but also evolve, highlighting new alternatives.

Our results show that both interacting Pairs and Individual participants were significantly affected by physical context, but Pairs were more likely than Individuals to use Abstract descriptions, irrespective of Layout. This difference in description scheme choice might be related to Pairs’ difficulty in ensuring a common interpretation of salient features of the maze when using Concrete description schemes, as the salience of these features can only be assumed to be shared by partners, and must be ratified in each description exchange. In contrast, Abstract schemes require no such repeated ratification once the system is shared by both speakers. Individual speakers, who do not receive external ratification of their descriptions, rely on the consistency of their descriptions to ensure comprehension by their future (imagined) interlocutor, and would therefore favour those description schemes that provided the best fit for their given layout.

### **3.4.3 Change from concrete to abstract description schemes: who, how and why?**

Our results showed that participants working in Pairs used more Abstract description schemes as they advanced through the game, even if they had not used Abstract descriptions in their initial exchanges. In particular, participants in Irregular layouts, who initially used more Concrete than Abstract descriptions, migrated towards Abstract descriptions after the first round; that is, they abandoned their original conventions for a new alternative.

As Garrod and Doherty (1994) predicted, the conflict arising from choosing to follow the precedent description scheme versus choosing to switch to a more salient scheme had the effect of transiently reducing the degree of inter-speaker coordination in interacting pairs. Overall, speakers' production of aligned or misaligned descriptions was mainly affected by round number, reflecting the fact that both Individual speakers and interacting Pairs became settled on a particular scheme after playing a few rounds. However, whereas this effect was equivalent for both Layouts in the Individual condition, Pairs results suggest participants in Regular layouts maintained higher alignment levels than participants in Irregular layouts. As Pairs in Regular mazes tended to maintain their original description schemes and therefore to align increasingly as they moved forward in the game, this pattern is consistent with Pairs in Irregular mazes tending to abandon their original schemes –and in this process locally misaligning with their partner– in order to subsequently converge again on a new scheme.

Two issues arise from the analysis of this migration: Why did participants abandon an established convention? And how was this movement performed in the context of dialogue? The interactive alignment account assumes that communication is successful if communicators come to understand relevant aspects of the world in the same way as each other (Garrod & Pickering, 2009). However, success in real life dialogue comes not only from a common understanding of the world, but also from coordination between the interacting partners and the world. An interacting couple could achieve perfect alignment in their conceptualisations, and at the same time completely fail in terms of the joint action being performed, in the same way that a pair of dancers could match their steps perfectly, and yet –if they failed to follow the tempo of the music– fail in the joint action of the dance itself. As in Mills (2014), our results show that the most highly coordinated dyads (measured in terms of the time they took to successfully solve a maze) did not necessarily use the same semantic model in the last turns of the game that they had used in the first turns. This flexibility implies that coordinated dyads are not mechanically repeating the schema used in their earliest interactions, but rather adapting to their changing circumstances, even if this means de-coupling temporarily. In this sense, a successful interaction will maintain coordination at a higher (task) level even at the expense of the lower (description) level (Garrod & Pickering, 2009).

In our experiment, the appropriateness or ‘fit’ of a description scheme with respect to a specific maze layout generated an initial contextual pressure, but the repetition of the task made some schemas more efficient than others over time, creating an additional and competing pressure. Pairs (but not Individuals) adapted to this new pressure by migrating towards a type of schema that could be more easily applied to new mazes or new positions in the same maze, namely Abstract schemas. In this sense, the language used in the Pairs adapted to its circumstances of use (Lupyan & Dale, 2016), where a system that facilitated repeated interaction was needed.

Why did Pairs show this adaptation over time, but Individuals did not? We propose that this adaptation was facilitated by the dynamics of dialogue, where a participant has the opportunity to evaluate the ease of comprehension and contextual applicability of their partner’s descriptions, and to apply that evaluation to their own descriptions if they correspond to the same schema, through pairing of comprehension and production (Pickering & Garrod, 2013). Thus dialogue affords an opportunity for ‘situated reflection’, whereby members of an interactive team can reflect on their own performance by observing their partner’s actions (Shirouzu, Miyake, & Masukawa, 2002). As Keysar & Henly (2002) noted, speakers usually overestimate the effectiveness of their own utterances; however, the comprehension process that occurs when participants evaluate their partner’s contribution can help facilitating the evaluation of the description scheme they themselves are using. Misyak and colleagues (2016) suggested that speakers engage in a *joint inference* process, where the consideration of both their own and their partner’s perspective allows them to jointly change their current conventions, dynamically adapting to the communicative and contextual needs of the pair. Crucially, this process would not be available for Individual speakers.

An alternative explanation for the shift from Concrete to Abstract schemes in terms of expertise seems less consistent with our findings. Under such an explanation, Concrete schemes might be considered as simpler, and Abstract schemes as more specialized and complex (Healey, 1997, 2008). The shift to Abstract schemes found in Pairs might thus arise from participants’ increasing expertise in the task of providing position descriptions. However, this explanation does not accord well with our overall pattern of results: Expertise can be accumulated by both Individual participants and Pairs, and so we would expect to see a similar switch to an alternative description



scheme in the Individual players, contrary to our findings. Moreover, a majority of the Pairs used Abstract description schemes from the first round, which contradicts the idea that Abstract schemes require some form of training. While it could be argued that Pairs had twice the number of opportunities for introducing change than Individuals<sup>7</sup>, the evidence suggest confirmatory feedback is crucial if a previously-used conceptualisation is to be dismissed. Even if the process by which an individual introduces a referential change could involve deliberate innovation (and in that sense be available for both Individual speakers and interactive Pairs), the crucial aspect of adaptation seems not to be innovation per se, but the possibility of abandoning a previously used conceptualisation, which is crucially made available by feedback.

The fact that Individual speakers showed little change throughout the game in terms of their position descriptions is compatible with more than one account. An alignment-based explanation would predict self-monitoring to act as a source of alignment within the speaker's own language processing system, as their linguistic output is examined for consistency between the intended output and the form actually produced. Individual speakers' consistency could also be explained in terms of cognitive economy, hence egocentric processing, as the introduction of a different perspective when this was not required by the task would certainly imply more processing effort. An audience design explanation (i.e., based on considerations of a partner's mental states) would similarly predict consistency, as the speakers' model of a future interlocutor cannot be contrasted with an actual co-present interlocutor, and is therefore fixed as an idealized model in their evaluation of the task.

Overall, our results are compatible with a 'language game' account (Wilkes-Gibbs and Clark, 1992; Brennan and Clark, 1996), where the particular history of interaction of each pair is considered. According to the language game framework, speakers go through a process of negotiation where a specific conceptual pact is agreed on, but which is provisional or open to change as the interaction unfolds. As such, references are chosen considering ahistorical factors (precedence, salience, lexical availability), but also adapt to changing (historical) circumstances (Brennan & Clark, 1996). Individuals, lacking interactive feedback, also lack an externally contrasted history

---

<sup>7</sup> However, Individual participants' results are not consistent with this view, as there is no pattern of change even in the third round of the experiment, while alignment is higher on each new round played, suggesting the maintenance of the same description schemes throughout the game.

which could feed their interpretation of the communicative needs they are aiming to fulfil.

#### **3.4.4 From pairs to communities**

The process by which a linguistic convention is established and perpetuated in time through dialogue can be linked to the process by which a language's lexical and syntactic elements are established and/or transformed over a historical timescale. However, the mechanisms by which change at the local interaction level affects (or is transformed into) change in the community at large, and from there to 'language' itself, are not obvious.

The conventions literature suggests that the selection of a linguistic variant in pairwise interaction is shaped mainly by the local effects of precedence and salience. But as our results suggest, the same processes are also open to the influence of the adaptation of that variant to the communicative task, which affects salience judgements by influencing the speakers' perception of their context: Speaking about a position as an X-Y coordinate highlights specific properties of the environment, whereas speaking about it as "*the very end of the tail*" highlights different properties. Adaptation to context could therefore link the selection of linguistic variants in pairwise interaction and their conventionalization in a community, by providing a stable pressure acting over different interactions and speakers.

The link between pairwise interactions and the community level of language change is directly addressed by Garrod and Doherty (1994), who argued (in line with Lewis, 1969/2002) that speakers turn to the schema they have most frequently encountered in the past when they face a new interaction. In this sense, the influence of their past interactions would be additive, with previously encountered variants competing with respect to the proportions of use in which they have been experienced. However, the description schemes emerging from the communities in Garrod and Doherty (1994) are not *conventional* in Lewis' terms, as there is no need for beliefs about the behaviour of other members of the community. As speakers do not need to be aware of the existence of a community at all to align on the most frequently encountered variant (though beliefs can modulate this tendency; see Branigan et al., 2011; Fehér,

Wonnacott, & Smith, 2016), Garrod and Doherty posit a more mechanistic explanation, where the global level emerges out of the sum of pairwise interactions.

Our results suggest, however, that alignment and frequency of use do not act alone in determining which variants will be adopted by a speaker, but that their content or 'value in context' also plays a role. Tamariz et al. (2014), using data from an artificial language experiment, explained the process by which these conventions form as a consequence of the interplay between two biases: the tendency to re-use an existing variant, and a preference for better (i.e. simpler or more communicatively functional) variants. The interplay between the two biases leads individuals to retain their own previously used variant unless they encounter a superior variant in terms of content<sup>8</sup>, in which case this new variant would be adopted. Under this account, communicative interaction, and in particular the evaluation of the communicative utility of alternative communicative tokens, is crucial to the spread of new variants and the formation of population-level conventions. Similar processes can be identified in our experiment. Participants in Pairs tended to align on the same description scheme variant, determined in the first place by the individual evaluation of the descriptions' properties against the context. However, as the interaction progressed, participants identified a superior variant in terms of its appropriateness in context and ease of use in repeated exchanges, and this variant was then adopted and stabilized in subsequent games.

### **3.5 Conclusions**

Alignment has been considered to be the main driver in the establishment of conventional references. However, as our results show, participants need to take into account more than just their partner's contributions when interacting in a given context. Both Individuals and Pairs of speakers produced references that were adapted to their context of use, but only interacting Pairs were able to dynamically

---

<sup>8</sup> The content value of an item in Tamariz et al.'s model is defined intrinsically, since the model does not include context as a factor; however, we can assume that, in real communication, the functionality of the item in context would affect its content value.

switch from an established reference to a new alternative when the task demanded a change. For alignment to act as a source of situated coordination, speakers need to consider both linguistic interaction and content functionality in order to reach an equilibrium that is not simple repetition of a partner's choices, but informed convergence.

### **3.6 Introduction**

In his time-honoured essay, David Lewis (1969/2002) argued that pairs of speakers cannot establish a fully-fledged convention, even if they consistently choose the same alternative and agree on their common use over time. In his proposal, the establishment of a linguistic convention requires a community of speakers, where even if an individual speaker does not abide to the convention, the belief that other members of the group will still hold it should make the convention prevail. By comparison, the behaviour of an interacting pair would be bound to the individual choices of the speakers, as each instance of use that differs from the local convention acts as precedent for a new alternative that could potentially replace the previous choice.

Garrod and Doherty (1994) showed experimentally how these differences modulated speakers' choice of description schemes in a maze game. In their experiments, communities showed ceiling levels of alignment in their reference choices after playing a few rounds of the game, reflecting the establishment of a convention. Pairs, on the other hand, were pervious to external factors, like the salience of a specific maze layout, that lead them to locally modify their choices. Even in the most aligned pairs, the local conventions established were vulnerable. Moreover, the experiments of Garrod and Doherty (1994) show that explicit beliefs about the existence of a community are not required for the establishment of linguistic conventions, as their participants were not aware that they were playing with members of the same group.

Considering this tendency to rapidly converge on a conventional choice, we might assume communities would be less inclined to switch to a different alternative, as a community speaker would not necessarily change their production when faced with a reference that deviates from the norm. However, we need to consider there is no previous research showing communities converging on anything other than the most abstract description scheme choices, therefore our experimental setup might not generate any convergence onto other schemes that would allow us to test this hypothesis.

In this pilot experiment, we applied the maze game paradigm described in Experiment 1 to small communities: participants, distributed into 8 4-person groups, played each round of the game with a different participant from the same group, creating a miniature network where all participants had interacted with each other by the end of the game. In all other aspects, the design and procedure remain the same as in the main experiment.

Even though the results included in this appendix should be comparable with those from Experiment 1, the low number of groups that could be tested determined that they were not included in the main body of the chapter.

## **3.7 Method**

### **3.7.1 Participants**

32 participants took part in the experiment, distributed into 8 4-person micro-communities. 18 participants were females and 14 were males; their mean age was 22.71 years; all were University of Edinburgh students who took part in the experiment for payment. All participants were native English speakers. Participants read and signed an informed consent form before taking part in the experiment.

### **3.7.2 Materials**

Materials were identical as in Experiment 1.

### **3.7.3 Design**

We used a 2x3 mixed design, with factors Layout (between-groups) and Round number (within-groups).

### **3.7.4 Procedure**

Participants came to the experimental lab in groups of 4 people. We ensured participants did not know each other in advance. They were given verbal and written instructions together, then seated in individual sound-proofed booths equipped with a network computer. They were informed they might play with either the same or a different partner in each round of the game. They were told they could communicate freely through the online chat tool, that they would exchange descriptions of the figures and their positions, and that they should aim to make the figure and its position identifiable to their current partner.

In each round, 2 interacting pairs were formed in each group, with the software connecting the two participants in each pair in one chat line. The chat box for each participant showed the number assigned to their current partner, and displayed their unfolding dialogue. At the end of each round, participants' screens showed a message asking them to wait while they were connected to their next partner. The overall design of the interaction sequence ensured that each participant played with a different partner from the same group in each round, thereby at the end of the game each participant had played with all other participants in their group (see Table 3.6). In all other aspects, task procedure remained the same as in Experiment 1.

Round	Pairs
1	1 & 2 - 3 & 4
2	1 & 3 - 2 & 4
3	1 & 4 - 2 & 3

**Table 3.6** Organisation of interacting pairs per round

### 3.7.5 Coding

Coding of participants' responses was similar to Experiment 1. Participants were assigned a group number according to the group they played at. Alignment was considered with respect to the descriptions of the two participants in direct interaction at each round of the game.

## 3.8 Results

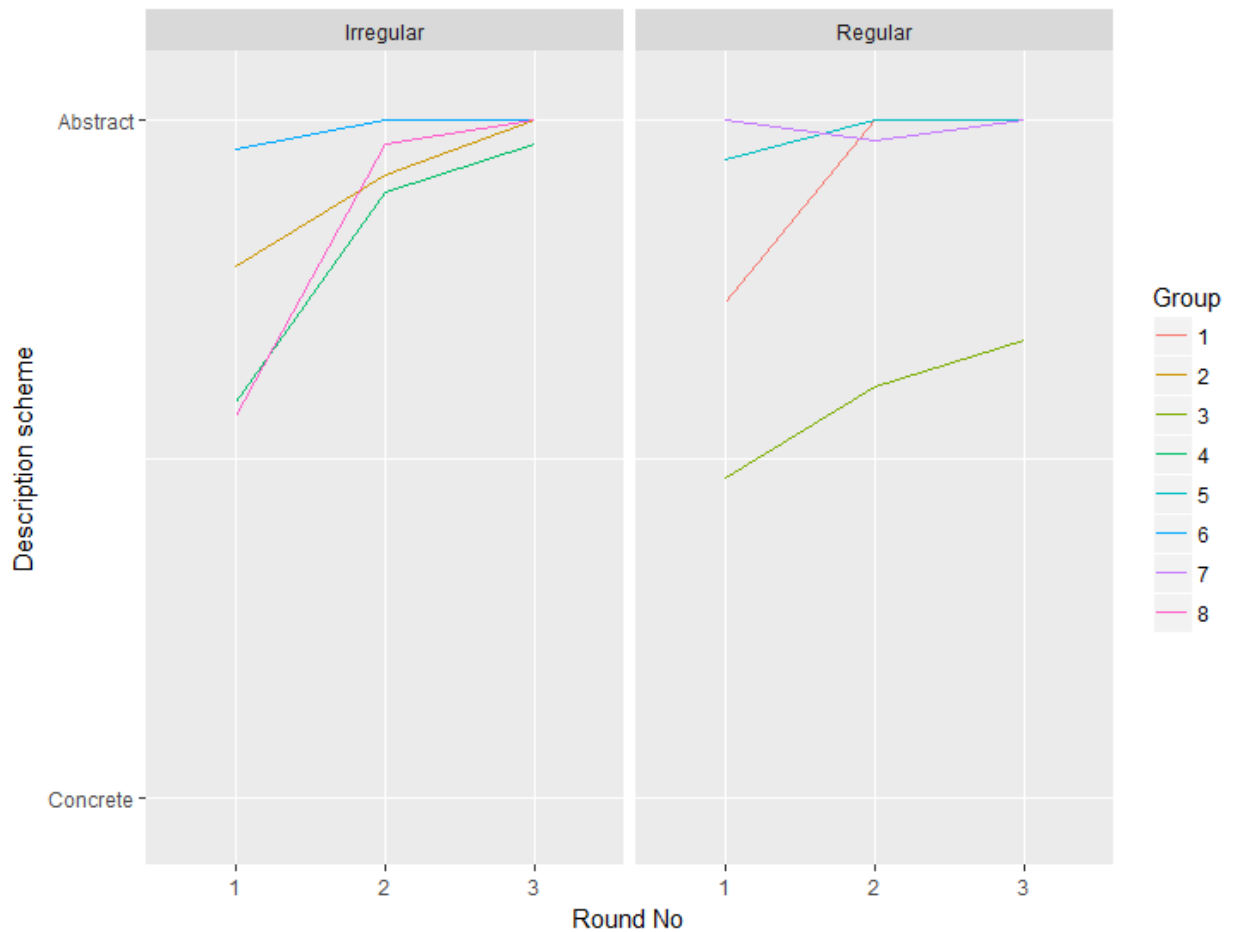
We used generalised linear models to estimate the effects of our experimental variables over participants' description scheme use. We fitted binomial mixed models with the variables Layout (factor, Irregular vs Regular) and Round (numeric, 1 to 3), and (unless otherwise specified) with a full random structure as per our design, including random intercepts for participant nested in Group, and a random slope for Group over Round number.

### *Choice of description scheme*

Using a full model with Layout (Irregular vs Regular) and Round number (1 to 3), plus their interaction, Round number had a significant effect over participants' use of Abstract or Concrete descriptions: participants used more Abstract descriptions with more rounds played ( $\chi^2=10.26$ ,  $p=0.001$ ). Layout did not have a significant effect over



participants use of Abstract or Concrete descriptions: Participants descriptions did not vary significantly whether they were playing on an Irregular or a Regular layout ( $\chi^2=0.63$ ,  $p=0.43$ ). The interaction between the two variables was similarly not significant ( $\chi^2=0.33$ ,  $p=0.56$ ).



**Fig 3.8** Mean description scheme use, per group. Facets show Irregular (left) and Regular (right) maze layouts. Description scheme is coded as Concrete (0) or Abstract (1), whereas mean per group is calculated as the numeric mean of all descriptions made by the participants in that group, per round.

**Table 3.7 Results of model of description scheme use**

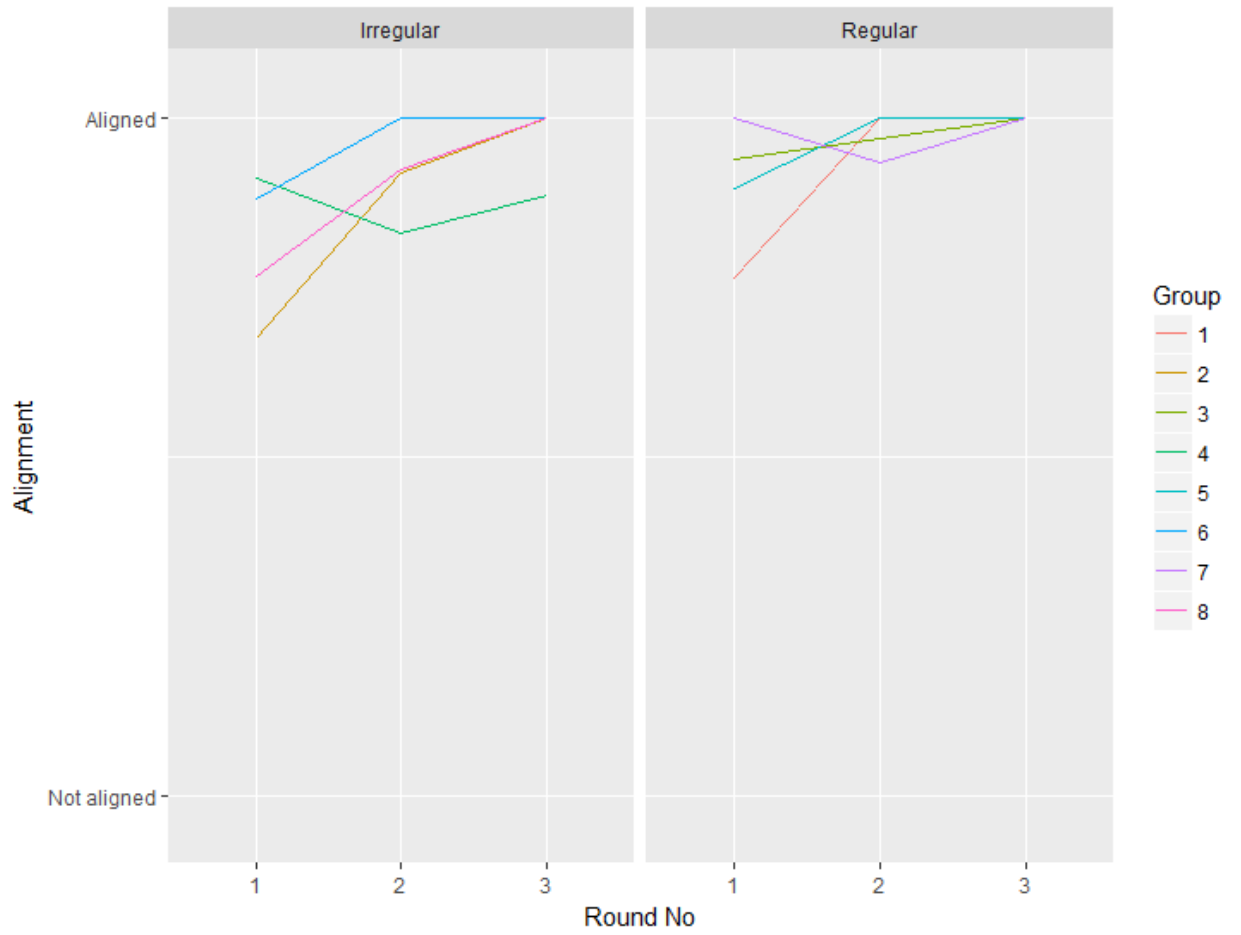
Predictor	Log-odds Betas	SE	X <sup>2</sup> and p values
Intercept	1.8772	0.5482	-
Layout	0.8502	1.0684	0.63, p=0.43
Round	2.3500	0.6858	10.26, p=0.001**
Layout * Round	-0.6987	1.1436	0.33, p=0.56

Log-odds betas, Standard Error, Chi-squared and P-values from Generalized Mixed Model fitted using *mixed* function from the *afex* package. Intercept represents grand mean at Round 1. Model is fitted with deviation-coded variable Layout (Irregular vs Regular). Round is coded as a numeric predictor, therefore model shows a linear effect of increasing 1 round.

### *Alignment*

Alignment was measured considering each description produced in a pair, and comparing it against the previous description by the other participant in the pair. Our model included Layout and Round, plus their interaction, as fixed effects, and a random intercept per participant nested in Group.

Model results show Round number had a significant effect over participants' alignment: Participants descriptions were more aligned as they played more rounds ( $\chi^2=40.28$ ,  $p<0.0001$ ). Maze Layout showed a marginally significant effect, with participants in Regular mazes using more aligned descriptions than participants in Irregular mazes. However, as Fig. 3.9 suggests, this marginal effect might be driven by one group.



**Fig 3.9** Mean alignment, per group. Facets show Irregular (left) and Regular (right) maze layouts. Alignment is coded as Not aligned (0) or Aligned (1), whereas mean per group is calculated as the numeric mean of all descriptions made by the participants in that group, per round.

**Table 3.8 Results of model of aligned descriptions**

Predictor	Log-odds Betas	SE	X <sup>2</sup> and p values
Intercept	1.9018	0.2175	-
Layout	0.7649	0.3978	3.73, p=0.05+
Round	1.4830	0.3120	40.28, p=<0.0001***
Layout * Round	0.8912	0.6234	2.41, p=0.12

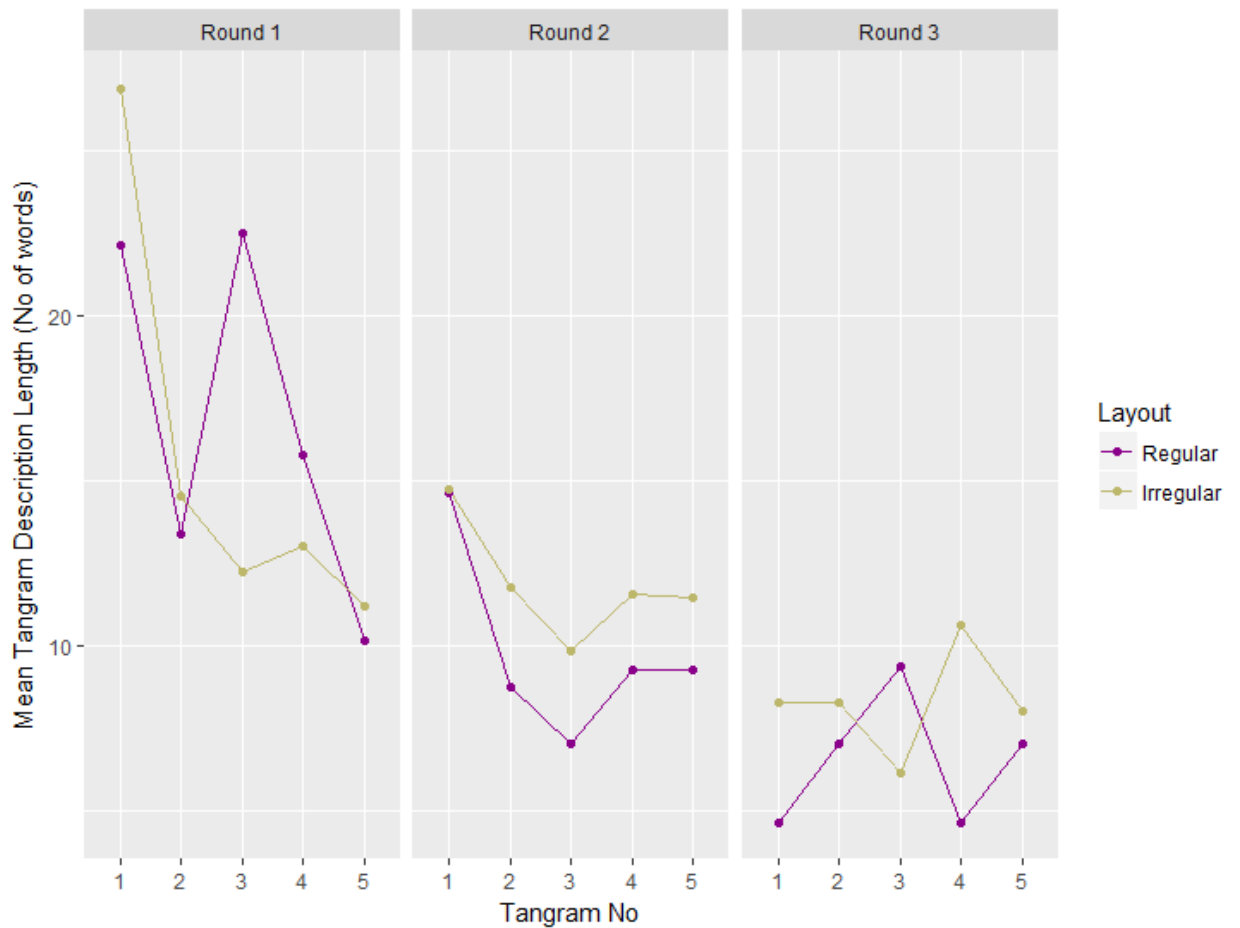
Log-odds betas, Standard Error, Chi-squared and P-values from Generalized Mixed Model fitted using *mixed* function from the *afex* package. Intercept represents grand mean at Round 1. Model is fitted with deviation-coded variable Layout (Irregular vs Regular). Round is coded as a numeric predictor, therefore model shows a linear effect of increasing 1 round.

### *Tangram descriptions*

Tangram description was measured in number of words per description, considering all utterances that referred to the same tangram in a round as pertaining the same description (therefore, a description could be co-constructed by both members of a pair, and usually it was). Within each round we measured the description length of the 1st to 5th tangram described; the 6th tangram was typically not described or referred to as “the last one”, since it was by that point the only remaining tangram.

We ran a linear mixed-effects model to predict the (log-transformed) number of words used per tangram description, with Round number, Layout, and Tangram number (1 to 5) and their interactions as fixed effect predictors. Random structure included a random intercept for Pair nested in Group, and a random intercept for Group (a random slope for Group in Round number was initially fitted, but discarded after the model results showed it to account for no variance and to be perfectly correlated with the intercept).

Tangram description length was found to decrease with Round number ( $\chi^2=41.91$ ,  $p<0.0001$ ) and Tangram number ( $\chi^2=14.39$ ,  $p=0.0002$ ): Participants' descriptions were shorter with each round played and each new tangram described. The interaction between both variables was also significant: Tangram description length had a stronger decrease with each new tangram described in Round 1, compared to Rounds 2 and 3 ( $\chi^2=10.75$ ,  $p=0.001$ ). There were no significant differences between Layouts ( $\chi^2=0.27$ ,  $p=0.60$ ).

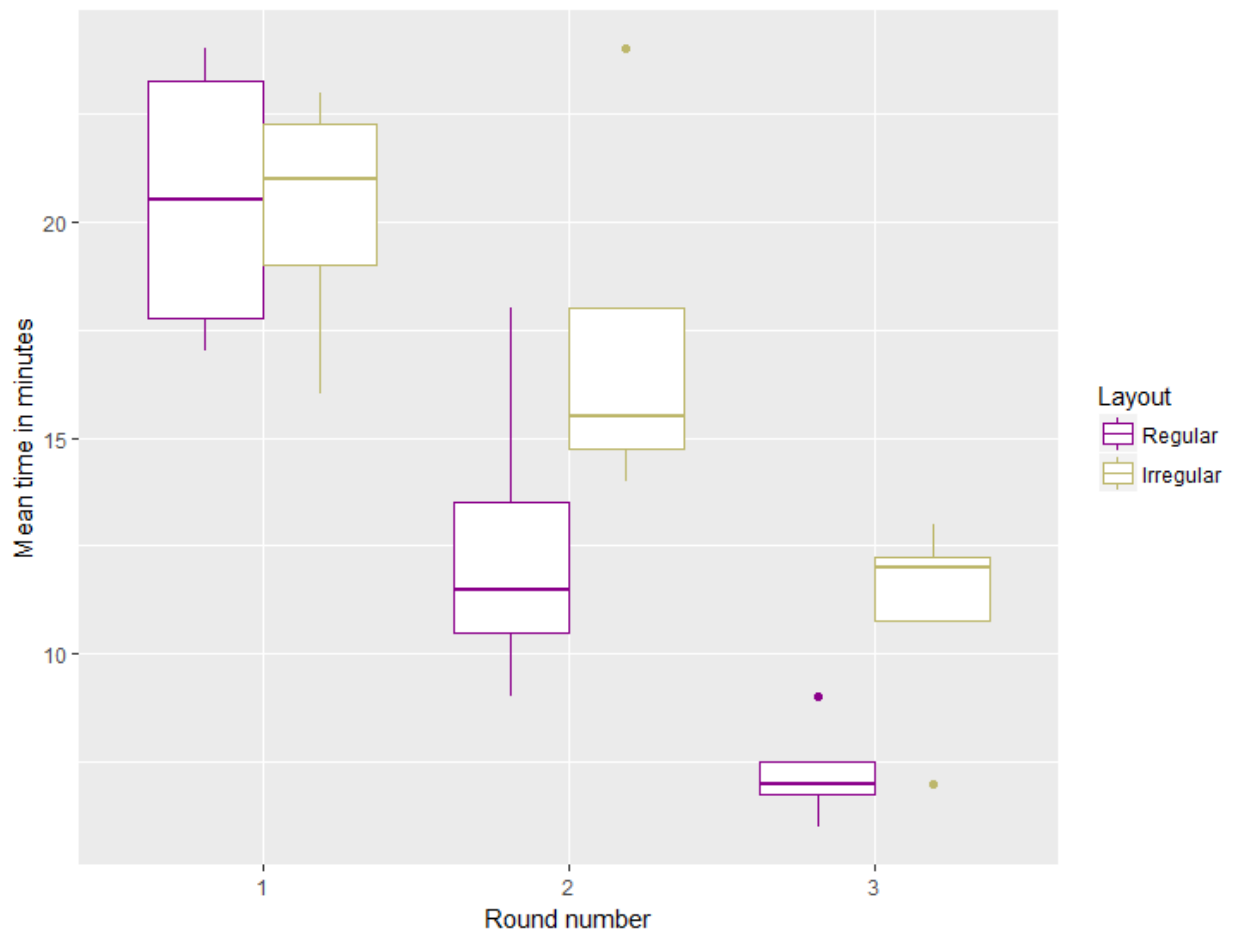


**Fig 3.10** Mean tangram description length, in number of words per round, across Layouts.

## Time

We modelled the total time each Group took to complete a round as a function of the Layout they were assigned to, their mean Abstract description use, and Round number, plus their interaction. Round number was the only significant predictor, with Groups taking an average of 6.4 minutes less to complete a round with a one unit increase in Round number ( $F(7,16)=7.592$ ,  $p=0.0004$ ,  $R^2=0.667$ ).

All Groups took significantly less time to complete a round as they advanced in the game, with no differences between Layout conditions nor Use of Abstract descriptions (but consider all groups showed high values of Abstract description use).



**Fig 3.11** Mean time taken per group to complete each round, per Layout.

### 3.9 Discussion

Overall, our experiment aimed at understanding the similarities and differences between the behaviour of individuals, pairs, and small communities of speakers with respect to their referential adaptation. In the main experiment we compared individuals and pairs, and in this appendix we included 4-people groups. Participants in these groups played one round with each of the other participants in their group, therefore forming an emergent community by the end of the third round.

The results showed high levels of Abstract description use from the first round and a steady increase per round, with a massive switch towards using exclusively Abstract descriptions by round 2 (all but one group showed more than 85% of Abstract description use on round 2). There were no significant differences between Layouts, with participants playing in Irregular and in Regular mazes showing similarly high levels of Abstract description use.

The high levels of Abstract description use across Layout conditions might be related to participants' evaluation of the concurrent demands of the task: on one hand, different groups of participants were facing different maze layouts; on the other hand, they were facing the common need to communicate with different people in each round of the game they played. Our results suggest the demands of the communication task were perceived as more pressing than the differential demands of the maze layouts, with participants across layouts choosing a more systematic approach to the task.

Participants were informed they might play with different people in each round before the beginning of the game, which might explain why groups in Irregular layouts show higher levels of Abstract description use in round 1, compared to Pairs and Individual participants. However, it was in round 2 when they had to face the need to communicate with a new participant after having established a system in round 1. In this situation, a description scheme that simplified the communication of positions in the maze by establishing a set of rules with an invariable interpretation might have appeared as the best alternative, explaining their massive switch to Abstract descriptions.

A comparison between the behaviour of Individuals, Pairs, and Groups, suggests that the need for communication, especially when it is actually demanded and not just imagined, appears as the biggest driver of referential adaptation. Speakers will aim for the reference that best fits their communicative needs, which can exert differential pressures over them depending on the nature of the communicative situation. An Individual only imagining a communicative situation might perceive this pressure to be minor, as she has no information that could guide her production other than her own evaluation of it. A Pair of interacting speakers might be guided by both their own needs and their partner's, slowly adapting their production to fit the communicative scenario; and a Group of speakers might be rapidly driven to find a system that allows them to effectively and efficiently communicate with different people.



## **Chapter 4. Drivers of reference change**

### **4.1 Introduction**

Most theories of language production agree in that speakers will tend to reuse the references that have been established during dialogue instead of switching to an alternative. In previous chapters, however, we have shown that context changes, lexical differentiation, or production needs might cue speakers to abandon previously used choices. In this chapter, we will report two experiments aimed at further investigating two suggested drivers of referential change: context changes, and participant role differences.

Experiment 1 showed how interacting speakers adapted to a demanding task by switching towards an alternative referential expression that would allow them to deal with the task in a more efficient way. In a maze game task that was repeated three times, interacting Pairs of participants used increasingly more abstract descriptions as they moved through the game, compared to Individual participants. However, the experiment also showed that, across Interaction conditions, participants were more likely to use Abstract descriptions as they advanced in the game, suggesting production experience affects the referential choices of not only interacting pairs, but also individual speakers.

Moreover, experimental research has pointed out that the rate of linguistic change in interacting pairs might be much higher than the rate of change in individual speakers, suggesting pairs would adapt faster than individuals (Smith & Wonnacott, 2010; Smith et al., 2013). Indeed, Experiment 1 showed that, while Pairs adapted at a higher rate, Individual speakers, who were producing descriptions for an imagined future participant, also tended to use more Abstract descriptions as they advanced in the task. Could it be that individual speakers would reach similar levels of adaptation in a more demanding environment? Specifically, would a changing

perceptual context push Individual participants into adaptation rates similar to those of interacting Pairs?

Experiment 2 was aimed at testing the effects of context change in the referential adaptation of Pairs and Individual speakers. We designed a task in which both Individual speakers, producing descriptions to an imagined future participant, and interacting Pairs of speakers were presented with an irregular maze, in which they would play the first 3 rounds of the game. After this, they would play 3 more rounds in either 1) the same irregular maze they saw in the first round, 2) a different irregular maze, or 3) a different regular maze. The experiment aimed at identifying whether a changing perceptual context would promote speakers' linguistic adaptation, comparing the influence of a maintained vs a changing context in the referential adaptation of Pairs and Individual speakers.

On the other hand, while our first experiment suggested interaction is one of the main drivers of linguistic adaptation, it did not address the question of what the specific features of interaction are that would facilitate the adaptation of referential choices to the communicative circumstances. In Experiment 3, we focused on the roles of the participant role of the interlocutors, and the previous interaction with a particular partner, as potential relevant factors in this adaptive benefit of interaction. Discussing the benefits of dialogue for comprehension, compared to monologue, Fox-Tree and Meyer (2008) suggested that the higher number of perspectives presented in most dialogues was key in producing output that was easier to understand. However, Branigan et al. (2011) contested this view and suggested grounded descriptions are easier to understand independently of the number of perspectives used in a conversation, pointing to a bigger role of feedback and addressee-centred design in creating references that are a better match for the communicative situation. Considering that perspectives that are easier to understand are probably well adapted to the communicative task they are aimed to fulfil, we can ask whether participants who take part in dialogue would be as likely to change their references to adapt to the demands of the task as participants who only witness the dialogue without taking part.

On the other hand, Brennan and Clark (1996) suggested interacting speakers that maintained the same partner after a context change were less likely to adapt to the

new context by abandoning their previously used (but overspecific in the new context) descriptions, compared to speakers that interacted with a new partner after the context change. This suggests speakers with a new partner should be more likely to abandon a previously used reference if they believe that a different alternative would be better suited to the task, as the link that was established between the old partner and the precedent should not necessarily be extended to the new partner<sup>1</sup>. Considering this, we can ask whether speaking to the same or a different partner would affect speakers' likelihood of switching to a more adaptive alternative.

Experiment 3 used a maze game task that presented participants with three different interaction conditions. In a Director-Matcher setup, participants played a first round as either Matchers or Overhearers, and 3 successive rounds as Directors with either the Director of the first round as Matcher, or a new participant. The experiment aimed at identifying 1) whether having access to different perspectives in dialogue is enough to yield an adaptive advantage, or whether direct involvement in the interaction is required, and 2) whether interacting with a different partner makes speakers more likely to switch to more adaptive referential choices than interacting with the same partner as previously.

Overall, this chapter will aim at clarifying the role of two of the most likely drivers of linguistic adaptation, context change and participant role differences, in the adaptation of speakers' referential choices.

## **4.2 Experiment 2: Relationship between context change and linguistic change**

Speakers use context as a background for communication in language production and comprehension. If two speakers are co-present, the fact that they can both attend to objects and situations in their common context is treated as a basis for reference, particularly if they can both assume the other person is actually attending

---

<sup>1</sup> The discussion whether these effects can be attributed to conceptual pacts being established between the speakers, or to memory cues linking the identity of the partner to the previous uses of the reference, has been hotly debated in the past decade (Metzing & Brennan, 2003; Horton & Gerrig, 2005; Horton, 2008; Brown-Schmidt, 2011).

to them (Clark & Marshall, 1981; Clark & Brennan, 1991). As such, context change has the potential of promoting referential change: from the strong pressure to specify a reference that comes from the addition of new elements in the same category than the target item (e.g. other shoes added to a context where just one shoe was present; Brennan & Clark, 1996), to the more subtle pressure of a change in the function of an item in context (Ibarra & Tanenhaus, 2016), the addition of new information to the conversation in the form of a context change might guide the abandonment of previous referential choices.

Experiment 1 showed how different perceptual contexts promoted the use of different description schemes that were initially better adapted to deal with salient features in these contexts. Regular maze layouts, which highlighted the maze's underlying grid structure, promoted the use of more abstract description schemes, while irregular maze layouts, with their indentations and protrusions, promoted the use of more concrete and idiosyncratic descriptions. We aimed to exploit this differential effect in the design of an experiment that would present participants with a context change. In Experiment 2, participants would begin playing in an irregular layout, which should promote the use of concrete conceptualisations, and would thereby be distributed into one of three conditions: A Same Maze condition, where they would continue playing in the same maze layout (acting as a no-change baseline), a Different Irregular condition, where they would switch to a different irregular maze layout, and a Different Regular condition, where they would switch to a regular maze layout.

Considering the results of Experiment 1, we expected participants in Pairs to increase their use of Abstract descriptions more than Individuals as they advance in the game, across Same Maze and Different Maze conditions, as interaction should promote higher levels of adaptation. Also, we expected participants in the two Different Maze conditions to increase their use of Abstract descriptions more than participants in the Same Maze condition, as exposure to different maze layouts should create a stronger demand for a description scheme that can adapt to these different contexts. Moreover, differences between the two Different conditions might also be noticeable, as a Regular maze layout should present a more explicit reference to the underlying grid structure that is common to all mazes, compared to a new Irregular layout. Additionally, the difference between Same and Different

Maze conditions might be greater in Individuals than in Pairs, considering Individuals should not be as moved to switch to a new alternative in a non-changing context (as the results of Experiment 1 showed), while Pairs might switch in both changing and non-changing conditions as a result of their interaction and the emergence of an additional pressure from the repetition of the task.

## **4.2.1 Method**

### **4.2.1.1 Participants**

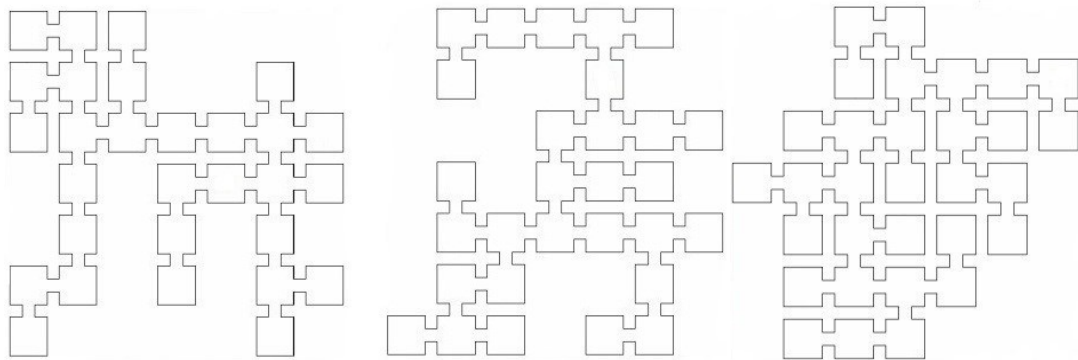
76 participants, comprising 26 Pairs of participants (39 females, 13 males; mean age 21.2 years) and 24 Individuals (18 females, 6 males; mean age 21.6 years), all University of Edinburgh students, took part in the experiment for payment. All participants were native English speakers. Participants read and signed an informed consent form before taking part in the experiment.

9 Pairs and 8 Individuals were assigned to the Same maze condition, 8 Pairs and 8 Individuals to the Different Irregular condition, and 9 Pairs and 8 Individuals to the Different Regular condition.

### **4.2.1.2 Materials**

We used two irregular mazes and one regular maze from the maze set created for experiment 1. As described in Chapter 3, maze regularity was calculated according to a neighbourhood-use metric, which defined regularity as the proportion of occupied vs empty space surrounding each maze square. According to this definition, a highly regular maze is a maze where most occupied squares are surrounded by other occupied squares, while an irregular maze is a maze where occupied squares are surrounded by a combination of other occupied squares and empty spaces. As such, an irregular maze is characterised by salient protrusions and indentations, while a regular maze reflects more closely the underlying grid structure.

One irregular maze (maze 1, on the left in Fig. 4.1) was used for the first 3 rounds, across conditions; the remaining 3 rounds were played in either the same irregular maze, the other irregular maze, or the regular maze, depending on condition.



**Fig 4.1** Mazes 1 and 2 (irregular, regularity scores 0.672 and 0.668) and 3 (regular, regularity score 0.864)

In each round of the game, the corresponding maze was presented, with 6 tangram figures randomly distributed across the maze squares. These tangram figures were the same as reported in experiment 1, chosen through an online survey to ensure that they were not too easily identified and not too difficult to describe (Fig. 4.2).



**Fig 4.2** Tangram set

Participants provided descriptions using a text-based chat tool (based on the DiET chattool software, Healey & Mills, submitted) which was configured to show one turn of dialogue at a time. This restriction was intended to roughly simulate the fading property of spoken dialogue, and to ensure participants would maintain a continuous interaction, instead of delivering all their information in one turn. Each participants' chat window displayed information about the typing status of the other participant (if available), remaining round time, and text of the current dialogue turn. Turns did not fade until the next turn by any participant was submitted.

The experiment software presented the mazes and tangrams and recorded participants' typed utterances, and was programmed using Java.

#### **4.2.1.3 Design**

Our design included 2 Interaction conditions: Individuals vs Pairs (between-participants); 3 Context conditions: Same Maze, Different Irregular, Different Regular (between-participants); and 6 rounds (within-participants). However, as the first 3 rounds were played in the same maze layout across participants, and only in the last 3 rounds were they split into the different Context conditions, most analyses were performed as 2 x 3 x 2, considering the 2 blocks of the game (rounds 1 to 3 vs 4 to 6).

#### **4.2.1.4 Procedure**

Participants came to the lab in pairs; we ensured that members of a pair did not know each other in advance. They were randomly assigned to Interaction and Context conditions (Individual vs. Pairs, Same maze vs Different Irregular vs Different Regular); they were informed of the Interaction condition they were to take part in, but not of the Context condition<sup>2</sup>. They were given verbal and written instructions together, then seated in individual sound-proofed booths equipped with a network computer. In the Individual condition, participants were told that they would be providing written descriptions for a future participant to follow, and that they should aim to make each figure and its position individually identifiable. In the Pairs condition, participants were told they could communicate freely through the online chat tool, that they would exchange descriptions of the figures and their positions, and that they should aim to make the figure and its position identifiable to their partner.

Participants were randomly assigned to one of three Context conditions: a Same Maze condition, where they played 6 rounds in the same irregular layout; a Different

---

<sup>2</sup> This implies that all differences in participants' performance per Context condition in rounds 1 to 3 are by definition random, as there was no difference in their tasks up to that point.

Irregular condition, where they played the first 3 rounds in an irregular maze layout, and the next 3 rounds in a different irregular layout; and a Different Regular condition, where they played the first 3 rounds in an irregular maze layout, and the next 3 rounds in a regular layout (see Fig. 4.1). All participants played a total of 6 rounds. Task procedure remained the same as in Experiment 1 in all other aspects.

#### **4.2.1.5 Coding of transcripts**

Participants' responses in terms of description scheme and alignment were coded in the same way as in Experiment 1.

#### **4.2.2 Results**

We will analyse participants' responses in two subsections: 1) Choice of description scheme (Abstract or Concrete descriptions), and 2) Alignment (use of the same description scheme as previous turn, per Pair/Individual).

##### **4.2.2.1 Choice of description scheme**

Our experiment presented Pairs and Individual participants with the same initial context: an irregular maze in which they had to describe specific locations for either a concurrent or an imaginary future partner to identify. All participants played three rounds in this irregular layout. In the second half of the game, participants switched or maintained their maze layout according to the Context condition they were assigned to: Same Maze, Different Irregular (a new irregular layout), or Different Regular (a new regular layout).

Descriptive statistics (Table 4.2, see also Figs. 4.3 and 4.4) show that participants across Interaction conditions and Contexts increased their use of Abstract descriptions from the first to the second halves of the game.

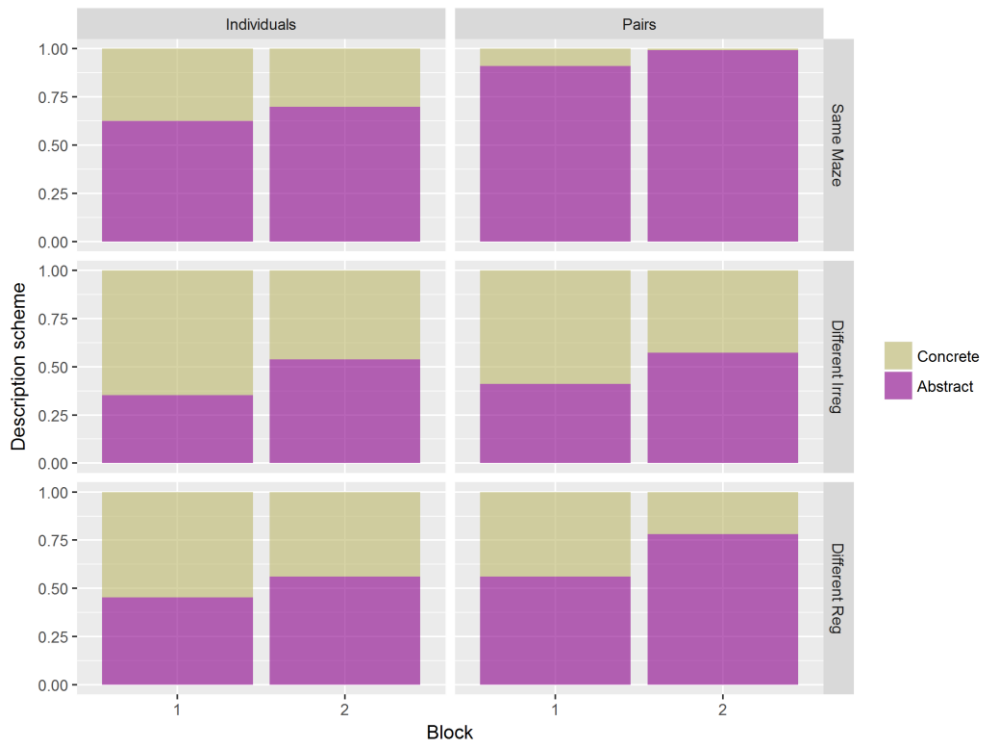


**Mean use of Abstract descriptions**

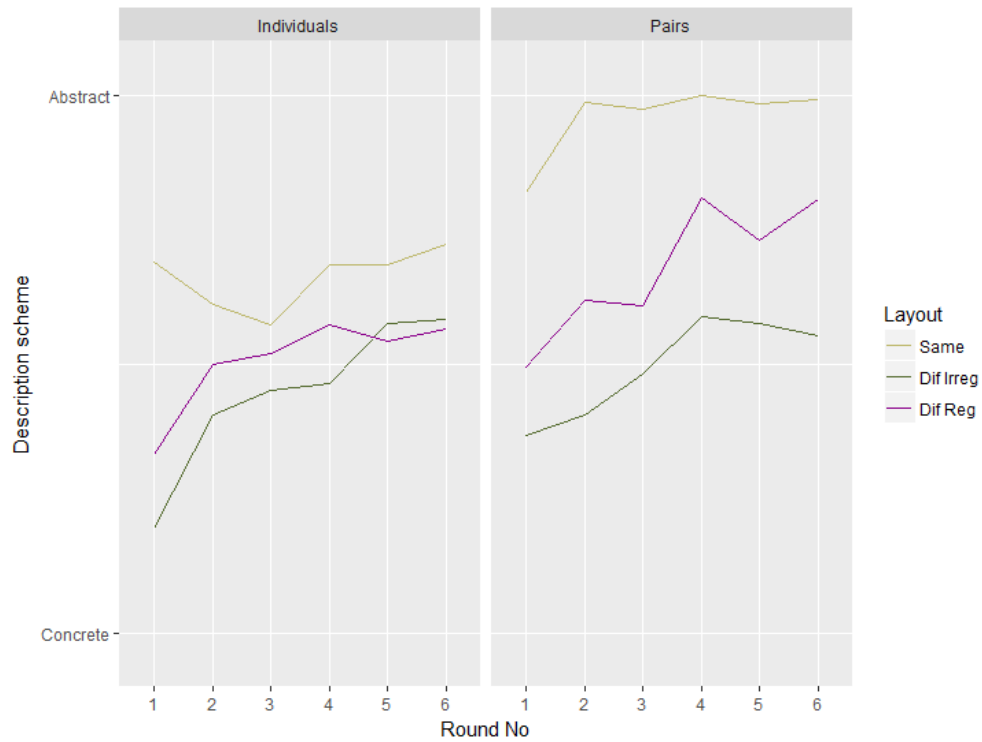
<i>Context</i>	<i>Interaction</i>	<i>Block</i>	<i>N</i>	<i>Abstract mean</i>	<i>SD</i>	<i>SE</i>	<i>95% CI</i>
Same maze	Individuals	1	163	<b>0.626</b>	0.485	0.038	0.075
Same maze	Individuals	2	162	<b>0.698</b>	0.461	0.036	0.071
Different Irreg	Individuals	1	125	<b>0.352</b>	0.48	0.043	0.085
Different Irreg	Individuals	2	117	<b>0.538</b>	0.501	0.046	0.092
Different Reg	Individuals	1	141	<b>0.454</b>	0.5	0.042	0.083
Different Reg	Individuals	2	141	<b>0.56</b>	0.498	0.042	0.083
Same maze	Pairs	1	479	<b>0.91</b>	0.286	0.013	0.026
Same maze	Pairs	2	372	<b>0.992</b>	0.09	0.005	0.009
Different Irreg	Pairs	1	553	<b>0.41</b>	0.492	0.021	0.041
Different Irreg	Pairs	2	430	<b>0.572</b>	0.495	0.024	0.047
Different Reg	Pairs	1	426	<b>0.561</b>	0.497	0.024	0.047
Different Reg	Pairs	2	358	<b>0.782</b>	0.413	0.022	0.043

**Table 4.1** Mean use of Abstract descriptions and associated descriptive statistics (description scheme use is coded as 0 (Concrete) and 1 (Abstract), numeric mean is derived from participants' total description use), per first and second blocks of the game (rounds 1 to 3 vs 4 to 6), per Context condition. N represents the total number of descriptions that were produced and coded in each condition.

We used logistic multilevel regressions (lme4 package in R, Bates et al., 2015) to analyse participants' use of description scheme according to the changes in their task context. Initially, we included Interaction (Individuals vs Pairs), Context (Same maze vs Different Irregular maze vs Different Regular maze) and Round number or Block (rounds 1 to 3 vs 4 to 6) as fixed effects, plus their interaction. The random structure was the maximal structure justified by our design (following Barr et al., 2013), including random intercepts for Pair/Individual and for Participant nested in Pair (to account for the initial variation between pairs and between subjects), and random slopes for Pair/Individual over Round number or Block (to account for the differential effect of task experience over the different Pairs or Individual participants). P-values were obtained through Likelihood Ratio Tests for each



**Fig 4.3** Proportion of use of Abstract vs Concrete descriptions, per Interaction condition and Context, in the first and second blocks of the game.



**Fig. 4.4** Numeric mean of use of Abstract description schemes per round. Description scheme use is coded as 0 (Concrete) and 1 (Abstract), line points represent mean per condition.

parameter against a model without that parameter, using the function *mixed* in the *afex* package (Singmann et al., 2017).

Using a full model with Interaction condition (Individuals vs Pairs), Context (Same maze vs Different Irregular vs Different Regular), and Block (First vs Second), plus their interaction, Block had a significant effect over the use of Abstract or Concrete descriptions, with participants using more Abstract descriptions in the second block vs the first block, across Interaction conditions and Contexts ( $\chi^2=12.81$ ,  $p=0.0003$ ): in the second block, the odds of participants using Abstract descriptions were 36% higher than in the first block<sup>3</sup>. We used deviation-coded contrasts to code for the different layout Contexts (Same maze considered as baseline), and compared participants' output in the first 3 rounds of the game (first block, same irregular maze across conditions) with their output in the last 3 rounds (second block, different mazes per condition) (see Table 4.2).

Neither the interactions, nor the main effects of Context and Interaction condition were significant: Participants' likelihood of using Abstract descriptions did not vary significantly depending on whether participants played in the same or a different maze ( $\chi^2=0.56$ ,  $p=0.45$ ), across blocks; nor between Individuals and Pairs ( $\chi^2=0.94$ ,  $p=0.33$ ).

Considering that the interaction between conditions was not significant in the full model, and the presence of an important numeric difference in the mean use of Abstract descriptions between Interaction conditions (particularly in the first block, see Table 4.1), we ran a second model without the interaction term. In this simpler model, both Different Context conditions (Different Irregular and Different Regular) were significantly different from the Same maze condition, considering the first block: Participants in the Different Irregular and Different Regular conditions were significantly less likely to use Abstract descriptions in the first block of the game, compared to participants in the Same maze condition ( $\chi^2=10.70$ ,  $p=0.005$ ). Block also had a significant effect: Participants in the Same maze condition were more likely to use Abstract descriptions in the second block of the game, compared to the first block ( $\chi^2=14.24$ ,  $p<0.001$ ).

---

<sup>3</sup> Participants' mean use of Abstract descriptions in block 1 vs block 2: 0.552 vs 0.690.

**Table 4.2 Results of model of description scheme use**

Predictor	Log-odds Betas	SE	X <sup>2</sup> and p values
Intercept	0.45144	0.77826	-
Interaction	1.51428	1.54959	0.94, p=0.33
Context	-1.44599	1.92482	0.56, p=0.45
Block	3.34482	0.92994	12.81, p=0.0003***
Interaction * Context	-0.09049	3.83053	0.00, p=0.98
Interaction * Block	1.64976	1.73112	1.02, p=0.31
Context * Block	-0.18740	1.96444	0.01, p=0.92
Interaction condition * Context * Block	3.23208	4.01971	0.69, p=0.41

Log-odds betas, Standard Error, Chi-squared and P-values from Generalized Mixed Model fitted using *mixed* function from the *afex* package. Intercept represents grand mean at First block. Model is fitted with deviation-coded variables Interaction (Individuals-Pairs) and Context (Same Maze vs Different Irregular vs Different Regular).

Overall, our results suggest participants across conditions increased their use of Abstract descriptions as they played more rounds of the game, but that this increase was not significantly different whether they were playing individually or in pairs, nor across the different layout contexts. Specifically, playing in the same or a different maze layout did not significantly modify participants' use of Abstract vs Concrete descriptions. Moreover, the presence of an initial difference in the use of Abstract descriptions that cannot be explained by our experimental manipulation needs to be further explored.

### *First round differences*

We ran an additional model to assess whether initial random differences between conditions might have influenced the pattern of results, using only data from the first round of the game (summary statistics in Table 4.3). The model included Interaction condition and Context as fixed effects, plus their interaction, along with a random intercept for Participant nested in Pair. The model revealed a significant effect of Context condition over the use of Abstract descriptions in the first round: Participants in the Same Maze condition were significantly more likely to use Abstract descriptions than participants in the Different Regular and Different Irregular conditions ( $\chi^2=8.90$ ,  $p=0.01$ ), across Interaction conditions. Neither Interaction condition ( $\chi^2=0.93$ ,  $p=0.33$ ) nor the interaction between Context and Interaction ( $\chi^2=0.28$ ,  $p=0.87$ ) were significant.

#### **Mean use of Abstract descriptions**

<i>Context</i>	<i>Interaction</i>	<i>mean</i>
Same maze	Individuals	0.702
Different Irregular	Individuals	0.176
Different Regular	Individuals	0.378
Same maze	Pairs	0.838
Different Irregular	Pairs	0.381
Different Regular	Pairs	0.5

**Table 4.3** Mean use of Abstract vs Concrete descriptions per Context condition, round 1.

#### **4.2.2.2 Alignment**

We used a logistic multilevel regression model to analyse participants' alignment in their descriptions according to the changes in their task Context and Interaction condition. We included Interaction (Individuals vs Pairs), Context (Same maze vs

Different Irregular maze vs Different Regular maze) and Block as fixed effects, plus their interaction. The random structure was the maximal structure justified by our design, including random intercepts for Pair/Individual and for Participant nested in Pair (to account for the initial variation between pairs and between subjects), and random slopes for Pair/Individual over Block (to account for the differential effect of task experience over the different Pairs or Individual participants) (Barr et al., 2013).

Using a full model with Interaction condition (Individuals vs Pairs), Context (Same maze vs Different Irregular vs Different Regular), and Block (First vs Second), plus their interaction, Block had a significant effect over participants' alignment in description use, with participants using more aligned descriptions in the second block vs the first block, across Interaction conditions and Contexts ( $\chi^2=16.37$ ,  $p<.0001$ ). The odds of participants using aligned descriptions were 3.6% higher in the second block of the game, compared to the first block (but consider high alignment scores in the first block, Table 4.5). We used deviation-coded contrasts to code for the different layout Contexts (Same maze considered as baseline), and compared participants' output in the first 3 rounds of the game (first block, same irregular maze across conditions) with their output in the last 3 rounds (second block, different mazes per condition). For participants in Pairs, alignment was coded considering their partner's last description; while for Individuals, it was coded considering their own previous description.

Our results suggest the Context manipulation did not affect participants' alignment, considering the first and the second blocks of the game ( $\chi^2=0.51$ ,  $p=0.47$ ). Interaction showed a significant main effect: Individuals were more aligned than Pairs throughout the game, across Contexts ( $\chi^2=7.12$ ,  $p=0.008$ ). However, this was to be expected due to the nature of the coding of their output, as each description was compared against the interlocutor's previous description in Pairs, and against the speaker's own previous description in Individuals.

**Table 4.4 Results of model of aligned descriptions**

Predictor	Log-odds Betas	SE	X <sup>2</sup> and p values
Intercept	3.2332	0.3393	-
Interaction	-1.5516	0.6030	7.12, p=0.008**
Context	-0.1981	0.7175	0.08, p=0.78
Block	3.1577	1.0773	16.37, p=<0.0001***
Interaction * Context	-0.3217	1.4309	0.05, p=0.82
Interaction * Block	1.0521	1.2412	0.65, p=0.42
Context * Block	0.8599	1.5136	0.33, p=0.57
Interaction condition * Context * Block	2.5591	3.0233	0.72, p=0.40

Log-odds betas, Standard Error, Chi-squared and P-values from Generalized Mixed Model fitted using *mixed* function from the *afex* package. Intercept represents grand mean at First block. Model is fitted with deviation-coded variables Interaction (Individuals-Pairs) and Context (Same Maze vs Different Irregular vs Different Regular).

### Mean use of Aligned descriptions

<i>Context</i>	<i>Block</i>	<i>Interaction</i>	<i>mean</i>
Same maze	1	Individuals	0.961
Same maze	2	Individuals	0.87
Different Irregular	1	Individuals	0.941
Different Irregular	2	Individuals	0.897
Different Regular	1	Individuals	0.932
Different Regular	2	Individuals	0.922
Same maze	1	Pairs	0.94
Same maze	2	Pairs	0.989
Different Irregular	1	Pairs	0.773
Different Irregular	2	Pairs	0.821
Different Regular	1	Pairs	0.833
Different Regular	2	Pairs	0.885

**Table 4.5** Mean use of aligned descriptions (description scheme use is coded as 0 (Not aligned) and 1 (Aligned), numeric mean is derived from participants' total description use), per first and second blocks of the game (rounds 1 to 3 vs 4 to 6), per Context condition.

### 4.2.3 Discussion

Context change has been identified as an important driver in reference adaptation and change (Brennan & Clark, 1996; Van Der Wege, 2009; Horton & Gerrig, 2002): even though speakers tend to reuse their reference choices, they are more likely to change them if a new context, new environmental features, or new goals, make the previous choice less appropriate. As Ibarra & Tanenhaus (2016) have suggested, a reference that has been used in interaction reflects a temporary agreement on how



to conceptualise an object in a particular context and with informational needs, and therefore it makes sense that the same conceptualisation will no longer be required if these context and informational needs change.

In this experiment, we asked participants to describe and navigate a maze game playing either in the same maze layout throughout the task, or in different mazes in the first and second blocks of the task. Our results suggest participants across Same Maze and Different Maze conditions used more Abstract descriptions as they advanced in the game, with a significant difference between the first and the second blocks of the game, across Interaction conditions. Our Context manipulation did not have a significant impact on participants' referential adaptation, with no significant differences in use of Abstract or Concrete descriptions for participants in the Same or Different Maze conditions.

While the Same Maze condition showed higher levels of Abstract description use, compared to the two Different Maze conditions (though not statistically significant), this difference might be derived from their higher use of Abstract descriptions in round 1 and across the first block of the game, a fact that cannot be explained by any experimental manipulation, as participants were given identical instructions and played otherwise identical tasks up to round 3. Moreover, these high initial levels of Abstract description use in the Same maze condition for both Individuals and Pairs of speakers might have influenced their adaptation patterns, especially in the case of the Pairs, who were using more than 80% of Abstract descriptions already in the first round.

As all Context conditions show similar levels of switch towards Abstract description schemes as participants advance in the game, we might tentatively conclude that the emerging pressure created by the repetition of the task appeared to the participants to be stronger than the pressure from the context change, causing them to use more Abstract descriptions that eased the effort of the task independently of whether they were playing in the same or a different context (though consider the initial difference in description scheme use).

Unlike in Experiment 1, in Experiment 2 both Individuals and Pairs of participants showed similar levels of switching towards Abstract descriptions. Two different

explanations might be suggested for this difference. On one hand, the fact that participants in Experiment 2 were aware they had to play 6 rounds of the task might have caused them to search for alternatives that could ease their effort earlier on in the game. This information might have provided an egocentric motivation for the switch that was not present in Experiment 1, where only 3 rounds were played. On the other hand, the impact that the context change had on participants' evaluation of the task might be easier to process individually than in pairs, as speakers in interaction would require to update their coordination procedures to the new circumstances. This would also explain the fact that Individuals in the Same Maze condition showed almost identical levels of Abstract description use in the first and second blocks of the game, while Individuals in the Different conditions showed a steep increase towards more Abstract descriptions as they played more rounds. Even though these differences are not statistically significant, they might suggest Individual speakers are more likely to adapt their references in a more challenging context, a point that remains to be further explored.

Overall, Experiment 2 did not provide significant evidence of an effect of a context change in speakers' switch towards different referential alternatives. Participants used more Abstract descriptions as they advanced in the game irrespective of whether they were playing in the same maze layout as before or a new layout, suggesting that the context changes provided by the different layouts were not considered by the participants to provide significant additional pressures. Specifically, the results suggest the emergent pressure derived from the repetition of the task had a strong influence on participants across conditions, overshadowing any minor impact the context change might have had on the adaptation process.

Further research is needed in order to understand how context change as an adaptation pressure would interact with other sources of reference determination. In particular, a more diverse task, where different referents need to be described, could help unravelling the effects of the repetition of the task and context changes, as the repetition of the description schemes in the maze game creates an experience effect that does not appear to be isolatable from other sources of adaptation pressure.

### 4.3 Experiment 3: Participant role in referential adaptation

Previous research has suggested pairs of interacting speakers are much more dynamic than individuals in terms of evaluating the effectiveness and efficiency of their references and adapting accordingly: they will eliminate superfluous elements from their descriptions (Krauss & Weinheimer, 1964, 1966; Clark & Wilkes-Gibbs, 1986); they will refine and abstract their graphical symbols (Garrod et al., 2007); and they will switch towards easier-to-use systems in repetitive tasks (Experiment 1).

At the same time, other research has suggested that, when speakers are asked to evaluate their own production, they tend to overestimate their effectiveness, implying they believe their references to be much more precise and appropriate than they actually are (Keysar & Henly, 2002). The overall picture suggests speakers gain a better perspective on their own production when interacting with a partner, and that the dynamic evaluation that is implied in each instance of the interaction allows them to act over this production, modifying what is needed to reach a balance between effort and effectiveness. However, interacting speakers are also driven towards alignment and conventionalisation, whereby an agreed-upon reference will continue to be used by both interlocutors, acting as a pairwise convention to designate a given item.

What is it about interaction that facilitates adaptation? Some accounts suggest pairing of production and comprehension could be part of the answer: if two speakers are aligned in interaction, they can evaluate the effect of their production choices as these same choices will be repeated back to them by their interlocutor, creating an opportunity for the comprehension, as addressees, of the same references they originally used as speakers (Pickering & Garrod, 2013). Other accounts have put emphasis on situated reflection: Interacting partners gain higher-level understanding of the task compared to individuals, as they get the opportunity to observe their partner's attempts at solving the task from a more detached, omniscient position (Shirouzu et al., 2002). In this account, the comprehension process is crucial, as it allows the speaker to understand the effect of the linguistic choices from the point of view of the addressee, which should allow her to choose the best alternatives for the communicative situation in her own turn to speak.

The interactive alignment account assumes communication is successful if communicators come to understand relevant aspects of the world in the same way as each other (Garrod & Pickering, 2009). So, if one speaker comes to 'see the world' in a different way, he should promote a change in the current convention which reflects this new conceptualisation. Considering that a pairwise convention will always be susceptible to change, as there is no community that would maintain the convention if an individual chooses a different option (Lewis, 1969/2002), the new perspective that the speaker has acquired should make its way into their discourse, potentially replacing their previous conceptualisation.

But do speakers need to engage in interaction to gain these adaptability benefits, or would witnessing an interaction be as beneficial? In other words, would exposure to a dialogue, including the multiple exchanges that might be needed to reach a common perspective in understanding, be as useful as actually taking part in the interaction to be able to adapt and switch to a 'better' conceptualisation? In an effort to understand what is it about dialogue that makes it 'better' than monologue in terms of comprehensibility, Fox-Tree and Mayer (2008; see also Fox-Tree, 1999) designed a task that presented participants with descriptions by either individual speakers or interacting pairs, crucially controlling for the number of perspectives that each fragment introduced (e.g. a fragment with one perspective would talk about a tangram conceptualised as "*a chicken*" either in a dialogue between two people or in a monologue; while a fragment with two perspectives would refer to it as either "*a chicken*" or "*a winged person*", either by two speakers in dialogue, or by the same speaker in different moments of a monologue). They found that the number of perspectives alone accounted for the benefit in comprehension, independent of whether the fragment came from one or two speakers. However, a similar study by Branigan et al. (2011) found that grounded descriptions (that is, descriptions that had been accepted by both speakers in a dialogue) were better in terms of comprehension even when they presented a single perspective, suggesting the interaction process that leads to grounding served as a kind of 'quality control' for the comprehensibility of these descriptions.

The identity of their interacting partner should also be relevant in the speaker's decision to maintain or abandon a referential precedent. Experimental work has shown that speakers are more likely to maintain sub-optimal references when

speaking to the same partner as before, compared to a new partner (Brennan & Clark, 1996). Similarly, Yoon and Brown-Schmidt (2014) showed that speakers were more likely to reconceptualise their descriptions (by adding new, different content words, compared to their previous descriptions) when addressing a new partner, compared to speakers addressing the same partner as before. Considering the likely role that the experience of their previous interaction played in the decision to modify their descriptions, we could also ask whether speakers would be more likely to adapt if they were talking to a new partner, compared to talking to the same partner as before.

We designed Experiment 3 with the aim of establishing whether having full access to a dialogue was as good as participating in it to drive adaptation to description schemes that were better fitted to solve a task<sup>4</sup>. The experiment further set out to establish whether speakers would be as likely to adapt and switch to a better perspective when speaking to the same partner as in previous rounds of the game, compared to speaking to a new partner.

Our task used the maze game in an atypical way: pairs of participants were distributed into Director or Matcher roles, with the Director guiding the Matcher to the positions of figures that were only visible in the Director's maze. After the first round, the roles would be reversed for the next 3 rounds, thereby controlling for who was the main person responsible for the production of position references in each round. Participants were divided into three Participant Role conditions: a Same Partner / Matcher condition, where the Matcher of the first round would play as Director in rounds 2 to 4, with the same partner they had in the first round as Matcher; a New Partner / Matcher condition, where the Matcher of the first round would play as Director in rounds 2 to 4, with a new participant as Matcher; and a New Partner / Overhearer condition, where an Overhearer in the first round (who had been 'witnessing' the interaction of an independent pair) would play as Director in rounds 2 to 4, with a new participant as Matcher.

---

<sup>4</sup> A caveat to this design is that overhearers have been found to have problems in understanding and interpreting referential descriptions (Schober & Clark, 1989). However, in our experiment the setup is highly similar for speakers in interaction and for participants observing the interaction, as the written chat medium prevents other non-linguistic factors from influencing understanding. Moreover, as the Overhearers only witness the interaction in the first round (without being asked to complete the task themselves), their ability to identify the exact positions described is not tested, allowing them to evaluate the speakers' use of different description schemes in more general terms.

We expected participants to be less likely to adapt their referential expressions if talking to the same partner, compared to a new partner, as both alignment and conventionalisation are more likely to occur if the same interacting partner is maintained. Moreover, we expected participants to adapt more when talking to a new partner if they had the opportunity to interact directly with a partner in the previous round, compared to only overhearing the interaction of other participants. If the grounding process is as important as the perspectives that are produced in an interaction for the generation of 'better' references, being a part of that process should bear a stronger benefit, in terms of the evaluation that the individual can make of the 'fit' between the description scheme that is being used, and the task. Overall, we expected participants in Condition 2 (New Partner / Matcher) to adapt more by switching to Abstract description schemes than participants in Conditions 1 and 3, and participants in Condition 3 (New Partner / Overhearer) to adapt more than participants in Condition 1 (Same Partner / Matcher).

### **4.3.1 Method**

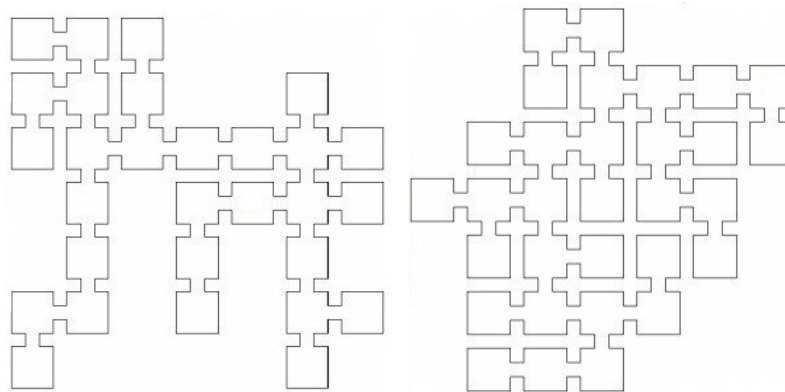
#### **4.3.1.1 Participants**

74 participants, comprising 43 females and 31 males (mean age 21.9 years), all University of Edinburgh students, took part in the experiment for payment. All participants were native English speakers. Participants read and signed an informed consent form before taking part in the experiment, and were given a debrief handout at the end.

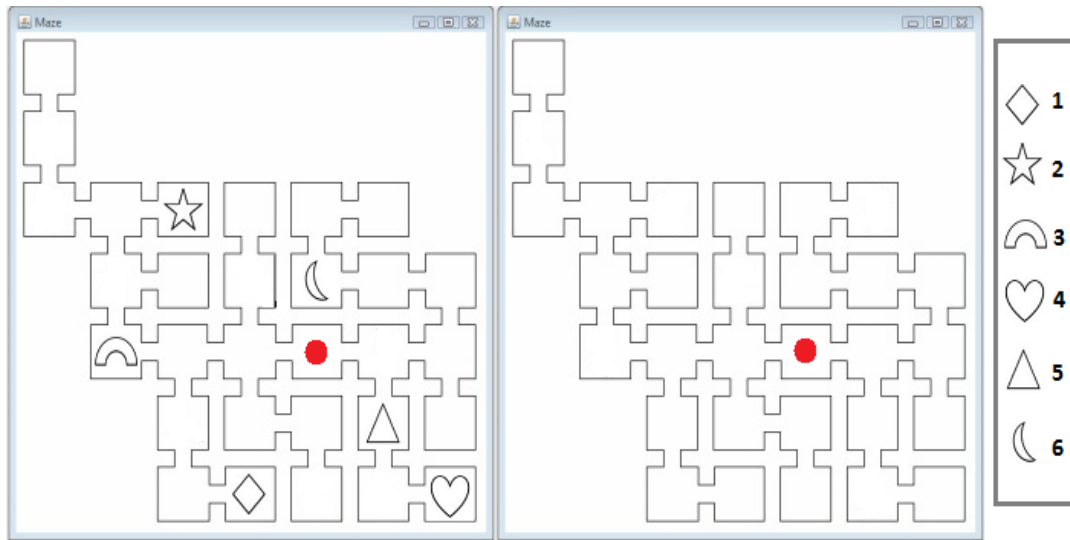
11 pairs of participants were assigned to the Same Partner / Matcher condition, 10 trios of participants were assigned to the New Partner / Matcher condition, and 11 pairs were assigned to the New Partner / Overhearer condition.

### 4.3.1.2 Materials

The experiment used two different mazes: an irregular maze, for round 1, and a regular maze, for rounds 2 to 4 (Fig. 4.5), both defined according to our regularity measure (described in section 2.1.2 and in Chapter 3). Both mazes had the same number of squares ( $\pm 1$ ). In each round of the game, the same maze was presented to both participants in a pair. The participant in the role of Director was presented with a maze that had 6 common geometric figures randomly distributed in the maze's available squares, while the participant in the Matcher role was presented with the same maze without any figures, and with a legend on the side of the maze that showed the 6 figures in a random order, numbered from 1 to 6 (Fig. 4.6). Participants in the Overhearer condition (round 1) received the same maze that was presented to the pair they were 'overhearing', showing the Matcher's configuration.



**Fig. 4.5** Irregular (left) and regular (right) mazes (regularity scores 0.672 and 0.864, respectively).



**Fig. 4.6** Regular maze (rounds 2 to 4) from Director (left) and Matcher (right) points of view.

Participants interacted through a text-based chat tool which was configured to show one turn of dialogue at a time. This restriction was intended to roughly simulate the fading property of spoken dialogue, and to ensure participants would maintain a continuous interaction, instead of delivering all their information in one turn. Each participants' chat window displayed information about the typing status of the other participant, remaining round time, and text of the current dialogue turn. Turns did not fade until the next turn by any participant was submitted. In the Overhearer setup, the same chat window delivered the transcript on the interaction as playback, maintaining the timing of the original dialogue. As the chat displayed the participant's name in each turn of the dialogue in all conditions, the output the Overhearer would see was very similar to the output each interacting participant received.

The experiment software presented the mazes and figures, and recorded participants' typed utterances and other keyboard activity. The experiment was programmed in Java.

#### **4.3.1.3 Design**

Our design included a Participant role condition: 1. Same Partner / Matcher, 2. New Partner / Matcher, and 3. New Partner / Overhearer (between-participants); and 4



rounds (within-participants). As the differences between Participant role conditions were apparent from the first round, all analyses are performed as a 3 (Participant role conditions) x 4 (rounds) model.

#### **4.3.1.4 Procedure**

Participants came to the lab in pairs or trios; we ensured that members of a pair or trio did not know each other in advance. They were randomly assigned to a Participant role condition (1. Same Partner / Matcher, 2. New Partner / Matcher, 3. New Partner / Overhearer). They were given verbal and written instructions together, explicitly stating which role each of them was going to be performing in each round of the game, and then seated in individual sound-proofed booths equipped with a network computer. They were told they could communicate freely through the online chat tool, and that their joint goal was to make the two mazes look exactly the same. To do this, the Matcher was required to provide the Director with the order of the figures, and the Director was to provide the Matcher with positions for those figures. The written instructions given to the Director included an example designed to prime participants to use Concrete descriptions, such as *“The heart is two up and two left from the starting point”*. This example was absent in the Matcher’s instructions; as such, the Director was directly primed to use Concrete descriptions, but the Matcher was only primed to use whichever description scheme the Director decided to use.

All participants played a total of 4 rounds. The first round was performed on an irregular maze. After completing the first round, participants would swap roles, with the Matcher of the first round becoming Director for the next 3 rounds, and either the Director of the first round (in Condition 1) or a new participant (Conditions 2 and 3) becoming Matcher (Fig. 4.7). The swapping procedure was performed automatically by the software.

In Condition 3, a participant would ‘witness’ the interaction of a previous pair in round 1 (randomly chosen from first round games in Conditions 1 and 2), by having the full transcript of their dialogue delivered one turn at a time in their chat window, while the corresponding Matcher’s maze configuration was presented in their

screen. After the transcript playback was completed, the participant would receive a new maze, becoming Director for the following 3 rounds, with a new participant (who had been waiting in a different booth) performing as Matcher.

Condition 1 (Same Partner / Matcher)	Condition 2 (New Partner / Matcher)	Condition 3 (New Partner / Overhearer)
A → B	A → B	A → B C
B → A	B → C	C → D

Fig. 4.7 Distribution of participants per Condition.

#### 4.3.1.5 Coding of transcripts

Participants' responses in terms of description scheme used and alignment were coded in the same way as in Experiments 1 and 2.

### 4.3.2 Results

We will analyse participants' responses in two subsections: 1) Choice of description scheme (use of Abstract or Concrete descriptions per dialogue turn), and 2) Alignment (use of the same description scheme as the previous turn, per Pair).

#### 4.3.2.1 Choice of description scheme

Our experiment presented participants with a variation of the traditional maze game task, where they were asked to perform different roles (Director, Matcher, or Overhearer) according to the Condition they were assigned to and the round they were playing. Summary statistics show participants used more Abstract descriptions as they played more rounds (Table 4.6, see also Fig. 4.8).

We used logistic multilevel regressions (lme4 package in R, Bates et al., 2015) to analyse participants' use of description scheme according to these changes in their participant roles. We included Participant Role condition and Round number as fixed effects, plus their interaction. The random structure was the maximal structure justified by our design (following Barr et al., 2013), including random intercepts for Pair and for Participant nested in Pair (to account for the initial variation between pairs and between subjects), and random slopes for Pair over Round number (to account for the differential effect of task experience over the different Pairs). P-values were obtained through likelihood ratio tests for each parameter against a model without that parameter, using the function *mixed* in the *afex* package (Singmann et al., 2017).

**Mean use of Abstract descriptions per round**

<i>Maze No</i>	<i>Condition</i>	<i>N</i>	<i>Abstract mean</i>	<i>sd</i>	<i>se</i>	<i>ci</i>
1	Matcher/Same Partner	148	0.311	0.464	0.038	0.075
1	Matcher/New Partner	74	0.514	0.503	0.058	0.117
1	Overhearer/New Partner	114	0.386	0.489	0.046	0.091
2	Matcher/Same Partner	123	0.642	0.481	0.043	0.086
2	Matcher/New Partner	64	0.688	0.467	0.058	0.117
2	Overhearer/New Partner	108	0.537	0.501	0.048	0.096
3	Matcher/Same Partner	98	0.745	0.438	0.044	0.088
3	Matcher/New Partner	49	0.857	0.354	0.051	0.102
3	Overhearer/New Partner	70	0.686	0.468	0.056	0.111
4	Matcher/Same Partner	92	0.652	0.479	0.05	0.099
4	Matcher/New Partner	41	0.854	0.358	0.056	0.113
4	Overhearer/New Partner	76	0.658	0.478	0.055	0.109

**Table 4.6** Mean use of Abstract descriptions and associated descriptive statistics (description scheme use is coded as 0 (Concrete) and 1 (Abstract), numeric mean is derived from participants' total description use), per first and second blocks of the game (rounds 1 to 3 vs 4 to 6), per Context condition. N represents the total number of descriptions that were produced and coded in each condition.

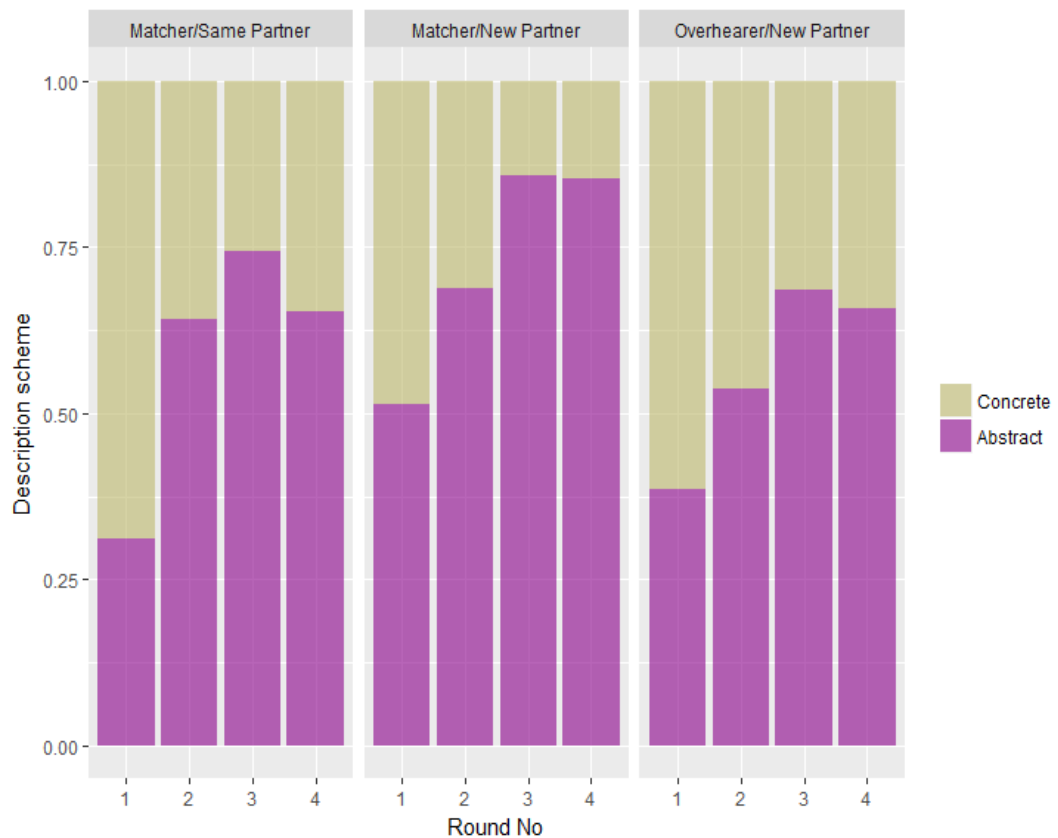
Using a full model with Participant Role condition (1. Same Partner / Matcher, 2. New Partner / Matcher, 3. New Partner / Overhearer), and Round (1 to 4, as numeric predictor), plus their interaction, Round had a significant effect over the use of Abstract or Concrete descriptions, with participants using more Abstract descriptions with an additional round played, across Participant Role conditions ( $\chi^2=9.47$ ,  $p=0.002$ ): with a 1 round increase, the odds of participants using Abstract descriptions increased by 24%.

Our results suggest the Participant Role manipulation did not affect participants' use of Abstract or Concrete descriptions: Participants use of Abstract or Concrete descriptions did not vary significantly according to the Participant Role they were assigned to ( $\chi^2=0.58$ ,  $p=0.45$ ). Specifically, use of Abstract or Concrete descriptions in Participant Role conditions 2 and 3 was not significantly different from condition 1 (we used deviation-coded contrasts to code for the different Participant Role conditions, with Same Partner / Matcher considered as baseline).

**Table 4.7 Results of model of description scheme use**

Predictor	Log-odds Betas	SE	X <sup>2</sup> and p values
Intercept	-3.5411	0.6058	-
Participant Role	1.1262	1.5005	0.58, p=0.45
Round	2.5073	0.7071	9.47, p=0.002**
Participant Role * Round	-0.1998	1.7753	0.01, p=0.91

Log-odds betas, Standard Error, Chi-squared and P-values from Generalized Mixed Model fitted using *mixed* function from the *afex* package. Intercept represents grand mean at Round 1. Model is fitted with deviation-coded variable Participant Role (Same Partner / Matcher vs New Partner / Matcher vs New Partner / Overhearer). Round is coded as a numeric predictor, therefore model shows a linear effect of increasing 1 round.



**Fig. 4.8** Proportion of description scheme used per round in each of the Participant Role conditions.

Even though the increase of Abstract description use was numerically smaller in condition 3, where the participant had only witnessed the interaction in round 1 without taking active part, this difference was not statistically significant.

#### 4.3.2.2 Alignment

We used a logistic multilevel regression model to analyse participants' alignment in their descriptions according to the different Participant Role condition they were assigned to, and the number of rounds they had played. We included Participant Role condition (1. Same Partner / Matcher, 2. New Partner / Matcher, 3. New Partner / Overhearer) and Round number (1 to 4, as a numeric predictor) as fixed effects, plus their interaction. The random structure was the maximal structure justified by our design, including random intercepts for Pair and for Participant nested in Pair (to account for the initial variation between pairs and between

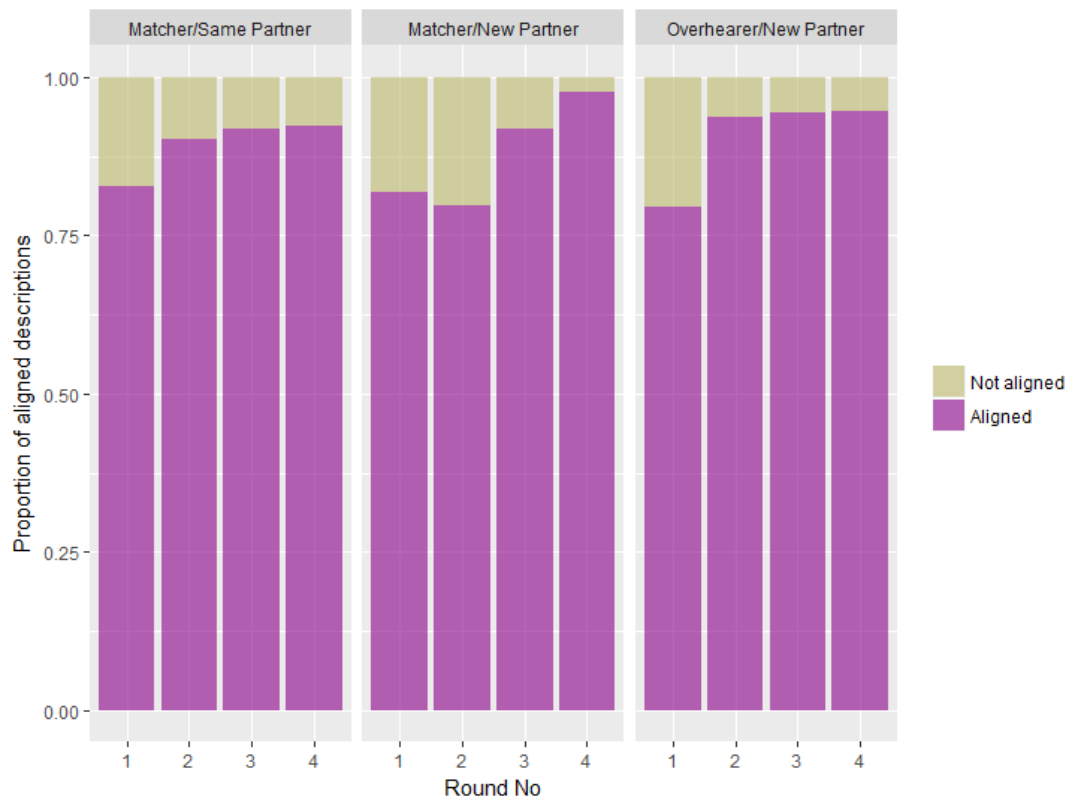
subjects), and random slopes for Pair over Round number (to account for the differential effect of task experience over the different Pairs of participants).

In a full model, which predicted Alignment scores from Participant Role condition and Round number, Round number had a significant effect over Alignment. Participants produced more aligned descriptions as they played more rounds, across Participant Role conditions ( $\chi^2=15.86$ ,  $p<.0001$ ): with a 1 round increase, the odds of participants aligning to the previous description increased by 20% (Table 4.8). Participant Role condition was not associated to significant differences in Alignment scores, with participants in the two New Partner conditions (conditions 2 and 3) not behaving significantly different from participants in the Same Partner condition ( $\chi^2=0.10$ ,  $p=0.75$ ). Proportions of aligned and not aligned descriptions are shown in Fig. 4.9.

**Table 4.8 Results of model of aligned descriptions**

Predictor	Log-odds Betas	SE	X <sup>2</sup> and p values
Intercept	0.6775	0.5730	-
Participant Role	0.4092	1.2866	0.10, p=0.75
Round	1.2109	0.3544	15.86, p=<0.0001***
Participant Role * Round	-0.6857	0.7506	0.88, p=0.35

Log-odds betas, Standard Error, Chi-squared and P-values from Generalized Mixed Model fitted using *mixed* function from the *afex* package. Intercept represents grand mean at Round 1. Model is fitted with deviation-coded variable Participant Role (Same Partner / Matcher vs New Partner / Matcher vs New Partner / Overhearer). Round is coded as a numeric predictor, therefore model shows a linear effect of increasing 1 round.



**Fig. 4.9** Proportion of aligned vs not aligned descriptions per round, in each Participant Role condition.

### 4.3.3 Discussion

Previous research suggests participant role might be relevant in an individual's commitment to conventional labels. While speakers and addressees usually converge on the use of the same references, which they hold as pairwise conventions, side-participants and overhearers might not be as committed to their use, as they didn't have the same opportunities to ground those references as the direct participants. We designed an experiment that set out to test whether speakers would show different levels of referential adaptation (the likelihood of abandoning a referential precedent in order to switch to a new reference that is better adapted to the task/context) according to their participant role in the task. Using a variant of the traditional maze game with defined Director and Matcher roles, we asked participants to play successive rounds of the game with either the same partner they had played in the previous round, or a new partner. An additional condition had

participants overhearing the first round of the game, in order to become direct participants in the following rounds.

Our results suggest our Participant Role manipulation was not strong enough to promote significant differences in speakers' referential adaptation. Even though the New Partner / Matcher condition showed higher levels of switching to Abstract descriptions, the differences between conditions were not significant. Round number, on the other hand, had a significant effect across Participant Role conditions, suggesting an incremental effect of experience dominated the adaptation process.

The numeric differences between our Participant Role conditions show a trend that partially follows the expected pattern of results. Condition 2 (New Partner / Matcher) shows the highest use of Abstract descriptions in rounds 2 to 4, in line with our prediction of more adaptation of speakers who had taken part in the dialogue and were interacting with a new partner. However, this difference might also be related to the higher initial levels of Abstract description use in Condition 2, which cannot be explained by task demands. As the proportion of Abstract descriptions used in a round is related to the proportion used in previous rounds, which grows as participants become more aligned throughout the task, a random initial difference is likely to have consequences in subsequent rounds.

Overall, alignment scores were high across all Participant Role conditions and rounds, which can be partially explained by our use of a Director/Matcher paradigm where only one individual was required to describe positions in each round. As such, the task produced dialogues that were less flexible in their use of references than previous maze game experiments. Moreover, these high levels of alignment do not seem to be reflecting a fixed description scheme choice, as participants use of Abstract description scheme varied across rounds. Instead, they suggest participants were much more likely to switch to a different description scheme between rounds, rather than within rounds. Even though we cannot statistically compare these results with those of previous experiments, this pattern suggests adaptation in this experiment was the result of a more global evaluation process of individuals in interaction, and less of a consequence of the subtle dynamics of natural dialogue.



#### 4.4 Conclusions

The experiments described in this chapter set out to analyse the effect of two potential drivers of referential adaptation, context changes and participant role, on speakers' change of reference choice. Experiment 2 used the maze game paradigm to present pairs and individual participants with either the same maze in each round, or different mazes in the first and second blocks of the game. Our results suggest this context manipulation was not enough to promote significant differences in participants' referential adaptation, with participants across conditions using more Abstract descriptions as they played more rounds. Contrary to previous research in referential adaptation, our experimental manipulation did not render the previously used references overspecific or sub-optimal, as all mazes could potentially be solved with either Concrete or Abstract descriptions. As participants switched to more appropriate references in all conditions, our results seem to suggest that a stronger experimental manipulation would be needed to discriminate between adaptation to context in different interaction roles, and simple accumulation of experience.

Experiment 3 used a variant of the maze game, where participants were assigned to a Director or Matcher role. Participants played a first round either as Matchers or Overhearers (witnessing the dialogue of a random independent pair), and the next 3 rounds as Directors with either the same partner as in round 1, or a new partner. Our results suggest the Participant Role manipulation was not enough to promote significant differences in participants' behaviour, with participants across Participant Role conditions increasing their use of Abstract descriptions as they played more rounds. As the numeric trends in our results seem to follow our predictions, with a more central role in the interaction (participant vs overhearer), and a new partner (as opposed to the same partner as before) both associated to more adaptation, we could assume a stronger manipulation might be able to determine whether different levels of involvement might impact speakers' referential adaptation. Specifically, the number of rounds participants played in either Matcher or Overhearer roles vs the rounds spent on Director role could be reversed, in order to create a task where the opportunity to change a referential schema comes after a longer history of differential involvement in the interaction.

Results from both experiments confirm participants are more likely to switch to Abstract descriptions as they play more rounds. Considering this effect has been replicated in most maze game studies (Garrod & Doherty, 1994; Healey, 1997; Mills, 2014), it can be assumed that any differences our experimental manipulations might have generated were masked by this overarching effect. Future studies might be able to determine whether context changes and participant involvement might cause significant differences in speakers' referential adaptation, considering a baseline drive towards adaptation that emerges as they gain experience in dealing with the task.

## **Chapter 5. Weighing of production factors in the use of referential precedents**

The experiments covered in this chapter will address the question of the interplay between egocentric and audience-based factors in the use of *linguistic precedents* – items that have already been used in the same communicative situation– in referential communication. As discussed in the previous chapters, even though in most dialogical settings speakers tend to repeat the linguistic choices they and their interlocutors have made in the past, there are several scenarios where they are likely to change these choices. Usually, these changes are produced to accommodate to a change in context, to respond to changing communicative needs, or to adapt to the requirements of the task (as in Experiment 1), but the influence of speaker's processing needs as drivers of these changes has not been thoroughly investigated. In an attempt to shed some light on this topic, we will report two experiments that address the influences of participant role and context over speakers' evaluation of their reference production.

### **5.1 Introduction**

The production of referential descriptions is a task speakers face on a daily basis. Even though extensive work has been carried out to explain the myriad factors that could have an influence in linguistic choices when producing a reference, the process by which speakers evaluate these different factors and weigh them in real-life dialogue is still unclear. Particularly, speakers' choice of either maintaining or abandoning a referential precedent –a description that has been used previously in the conversation– has mostly been looked at from a time-course perspective, where researchers have tried to determine which factors are available at which point in processing. However, even though researchers disagree on the availability of

audience-centred information at early stages in production, most studies agree with a role for some kind of model of the addressee's perspective in the speaker's processing of references, which is consistent with the evidence from natural language use in conversation (Horton & Gerrig, 2005; Arnold, 2008). In considering their addressee's perspective, speakers need to evaluate the information they share and don't share with them and how certain they are of this fact, in order to make a choice that is not only safe (to avoid having to incur additional turns or repair) but also efficient in terms of production resources. In the sections that follow we will review the role of egocentric and audience-centred perspectives on production and the factors that influence lexical choice, particularly common ground. Additionally, we will analyse the status of referential precedents from the speaker's perspective, looking at how differences in the addressee's participant role might affect their certainty.

### **5.1.1 Self and other(s) in reference choice**

The use of egocentric and audience-centred information in speakers' choice of references has been hotly debated since the publication of H.H. Clark and colleagues' classic works on common ground (Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989; Clark & Brennan, 1991; Wilkes-Gibbs & Clark, 1992). The authors proposed that speakers keep track of the information they mutually share with their interlocutor(s) and use this information to determine when to use a linguistic precedent and when new information is required. An opposing perspective was put forward by Keysar and colleagues (the 'monitoring and adjustment model' of Horton & Keysar, 1996; Keysar, 1997; Keysar, Barr, & Horton, 1998; Keysar, Barr, Balin, & Paek, 1998; Barr & Keysar, 2002; Keysar & Barr, 2005), who argued common ground is only accessed by the speakers as a secondary source of information, mainly to diagnose and correct coordination problems, with the bulk of the processing employing an egocentric perspective (see also Bard & Aylett, 2005). A third perspective proposed that what is traditionally thought of as common ground use in language processing can be better understood as the effects of general memory processes shaping the information available to speakers in dialogue, a proposal that does not require any appeals to guided perspective processing (Horton & Gerrig, 2005, 2016).

Hanna and Tanenhaus (2005) have more recently argued against a purely egocentric account, showing experimentally that the speaker's perspective can be considered by the addressee from the very first moments of comprehension in object disambiguation tasks. In their work, participants are able to constrain the possible interpretations of ambiguous utterances by considering the perspective of the speaker, to whom only one interpretation is available. Additional work by Hanna and colleagues (Hanna, Tanenhaus, & Trueswell, 2003) has argued that the use of common ground or egocentric perspectives is flexible and probabilistic, a position that could explain the different patterns of results found by different research groups.

To this point, most of the experimental work designed to discriminate between a main egocentric or other-centred influence on maintaining or changing referential choices has dealt with the consequences for listeners, who might or might not benefit from a *same speaker advantage for maintained precedents*, or a *different speaker advantage for broken precedents* (Metzing & Brennan, 2003; Kronmüller & Barr, 2015). From the speaker's perspective, the interplay of these two factors has been studied looking at production choices, mostly addressing the differential production of appropriate vs under- or over-specified descriptions (Horton & Keysar, 1996; Brennan & Clark, 1996).

While Barr and Keysar (2002) showed that listeners expect the same labels to be maintained when referring to the same objects, as Brennan and Clark (1996) had predicted, it is not clear whether speakers would maintain overspecific labels due to their consideration of the listener's expectations, or rather due to ease of processing of primed labels. Moreover, there is no evidence that speakers are aware of this potential benefit for the addressee in maintaining the same labels, nor of any benefit for themselves as a consequence of the benefits for the addressee. Further work by Galati and Brennan (2010) found that speakers attenuated the information they gave on a narration (using a lower amount of detail in descriptions of events and characters, and lower clarity of articulation), thereby lowering their own production effort, if talking to the same addressee for a second time, compared to talking to a new partner, suggesting some degree of audience-design, but also an egocentric motivation. Taken together, the evidence seems to suggest speakers consider preferentially their own needs, but that they are aware of the needs of their partners.

The question is, then, what does it take for them to take their partners' needs into account.

Regarding the commonality of precedents, work by Shintel and Keysar (2007) asked whether it was relevant for listeners to know that a precedent had been mutually established (in a *conceptual pact* fashion, that is, accepted in interaction by both speakers) to expect speakers to maintain it, and found that listeners expected precedents to be maintained independently of whether they were mutually known. Precedents serve as linguistic indices to the representation of the referent in memory, and that link, presumably faster and easier to retrieve than interpreting a full descriptive reference against context, should be highlighted in both speakers' and listeners' memories (Barr & Keysar, 2002). Following this question, we could ask whether it is necessary for speakers to know that a precedent has been mutually established to maintain it, or whether it is only necessary that they themselves used it (as recent work by Fukumura and Van Gompel (2012) seems to suggest). Section 5.1.3 will discuss these possibilities.

### **5.1.2 Lexical choice and overspecification in dialogue**

When a speaker faces the task of providing a referential description for an addressee to identify this referent in a given context, several competing forces are likely to be involved. Firstly, the words she might choose appear with different frequencies in her language, so their availability in her mental lexicon will differ. Secondly, her own previous use of language might be influencing the activation of these words in her memory, causing some words to be more readily available for use in production or comprehension (Barr & Keysar, 2002). Thirdly, the context might make some word choices more or less likely to be used, considering a myriad of contextual factors like the addressee's identity and their shared history, the expertise of the addressee in the topic, other items in context that might complicate target identification, cultural/group preferences, and so on.

For most objects, the most readily available label should be the basic-level noun that uniquely defines it as a unit, such as *chair*, *car*, *bird*, etc., as opposed to superordinate-level nouns such as *animal* or *furniture*, and to specific, subordinate-

level nouns such as *convertible* or *sparrow* (Rosch, Mervis, Gray, Johnson, and Boyes-Braem, 1976). According to Cruse (1977), the most common term that is informative enough to uniquely identify the referent should be considered unmarked, while other terms should be considered marked, and therefore their use explained by other properties of the context. However, this does not consider how experience might modulate the ‘default’ concept that is used by a speaker. Tanaka and Taylor (1991) showed how expertise on a topic affected the use of basic level labels, with experts on a given field using more specific labels to refer to items in that field (e.g. a bird-watching enthusiast using “*starling*” where most speakers would have used “*bird*”). We can assume that interaction and the speaker’s history of use of a label might similarly influence their label choice, with speakers being more likely to use non-basic level words depending on their cumulative history of interaction regarding that word and topic.

In the course of a dialogue, the history of interaction between the interlocutors adds other sources of information to the lexical selection process. Lexical entrainment explains the repetition of the same word choices to refer to the same items. Lexical precedents are established as the speakers use specific words for specific referents, and conceptual pacts can be established as interlocutors agree to the use of a specific conceptualisation to refer to an object. As priming and frequency effects can affect lexical availability *in general* (Bates, Masling, & Kintsch, 1978), lexical precedents affect availability for that particular conversation, and *for those particular interlocutors* –though some discussion exists on how exactly this is to be understood, as described in the first part of this chapter. Moreover, reference-meaning pairs that have been repeatedly used by a pair of interacting speakers become *routinized* for the purposes of their interaction, as the links between the words and the specific meaning they have been used to mean become stronger and more directly accessible (Pickering & Garrod, 2004; Garrod, 2011).

According to Grice’s (1975) Maxim of Quantity, speakers should provide their interlocutor with as much information as needed to successfully identify their intended referent, but not more than is needed. In the absence of competitor objects, the reference to a common item will then likely correspond to the ‘unmarked’ basic-level noun, in Cruse’s terms (e.g. “*that’s my dog*”), while the use of over- or underspecific information should be taken by the listener as an indicator of

a different meaning being implied (e.g. “*that’s my Scottish terrier*” signalling that I have other dogs, or “*that’s my pet*” highlighting my specific relationship to it). However, overspecification, defined as a speaker including more information than what would be needed to uniquely identify the referent but without a different communicative intention, is common in natural language use.

Colour has been identified as the feature of objects most likely to be overspecified in descriptive statements (Tarenskeen et al., 2015; see also Sedivy, 1999, 2005). According to Rubio-Fernández (2016), the mention of the colour of an object, even when not necessary to establish a unique reference, might help facilitate object identification, as it usually rests on perceptual salience and contrast in context<sup>1</sup>. In spontaneous mention, Sedivy (1999) found that participants produced unnecessary colour adjectives in descriptive statements nearly half of the time, compared to scalar adjectives that were rarely produced when not required for contrasting purposes. Overspecification is also susceptible to priming, with speakers using more overspecific references after hearing or using an overspecific reference (particularly if the syntactic structure of the references is similar, and if the dimension being overspecified is the same) (Goudbeek & Krahmer, 2012).

Even though in most contexts overspecification can be assumed not to be costly to the listener (though in specific contexts it can cause comprehension problems or delays, as in Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Engelhardt et al. 2011), speakers usually avoid producing redundant information, as the cost of production would be increased without a communicative payoff. However, if a specific description, like “*the Scottish terrier*” is used in dialogue in an appropriate context (e.g. where other dogs are present, rendering a basic-level label ambiguous in context), what should a speaker do when the context changes (i.e. no other dogs are present) and the use of a specific description is no longer justified? The speaker is faced with competing demands: on the one hand, she should avoid producing more information than is needed to uniquely identify the referent, and on the other

---

<sup>1</sup> Even though Engelhardt et al. (2011) found a detrimental effect for the comprehension of overspecific descriptions compared to noun-only descriptions, their results show that colour overspecification elicited reaction times that were less than half those of size overspecification, suggesting that even in contexts where more information could be detrimental, colour is much more easily processed than other dimensions. Moreover, their main pattern of results might be partially driven by their use of simple 2-object displays, where object identification is already at its lowest difficulty level (Rubio-Fernández, 2016).



hand, she should maintain the lexical precedent that was established with her interlocutor in the initial descriptions of the item. This was the question posed by Brennan and Clark (1996), who showed that a majority of speakers continued to use entrained descriptions even when these became overspecific for the current context<sup>2</sup>. Additionally, they found a difference in speaker behaviour according to the identity of their addressee: speakers continued using overspecific descriptions if they continued interacting with the same partner, but switched to basic-level labels if they were interacting with a new partner.

Brennan and Clark discussed their results in terms of conceptual pacts: a speaker will maintain an overspecific description if they continue speaking to the same partner because they share a conceptual pact with them, and changing the conceptualisation they agreed on would signal to the partner that a change of meaning is implied (E. Clark, 1997). Inversely, they would abandon the overspecific descriptions if the partner with whom the conceptual pact was established is no longer present. However, even if no conceptual pact is established, speakers establish lexical precedents when they use a conceptualisation, such that they are primed to repeat the same conceptualisation when they are faced with the same referent. This fact might explain why more than half of Brennan and Clark's participants retained their entrained overspecific descriptions even when they were speaking to a new partner, and even after 3 rounds of playing in the non-specific context.

### **5.1.3 What does it take to form a conceptual pact?**

Even though related, the terms *common ground* (Stalnaker, 1978; Clark, 1996) and *conceptual pact* (Brennan & Clark, 1996) are fundamentally different. While *common ground* refers to all the information that is common to both interlocutors and believed by them to be so, including the knowledge of the linguistic precedents they have used during the current conversation and previous exchanges, a *conceptual pact* implies more than mere adherence to a previously used linguistic precedent. It reflects the idea that a conceptualisation in dialogue, whether it is actively used by

---

<sup>2</sup> Complementary, Barr and Keysar (2002) found that listeners expected speakers to maintain referential precedents even when these became overspecific in a new context.

both speakers or only used by one and ratified (by action or omission) by the other, establishes an agreement between them that states that *this*, and not any other, is the conceptualisation that will be used to designate that item.

Extending the classic example, if two speakers are discussing opposing views on abortion, with one speaking about “*the foetus*” and the other calling it “*the baby*”, the fact that they have used both conceptualisations is common ground between them, but no conceptual pact has been established. The mere belief that a conceptualisation has been used and understood does not establish a conceptual pact, as there has to be some sort of ratification of the acceptance of that specific conceptualisation by both interlocutors.

Both everyday conversations and experimental scenarios like tangram games offer examples of this divergence. In dialogue, the fact that a specific perspective on an object has been understood does not necessarily mean that a conceptual pact has been established, as new subsequent perspectives might be brought to the conversation. Consider the following example (taken from Clark & Wilkes-Gibbs, 1986):

- A. *Okay, and the next one is the person that looks like they're carrying something and it's sticking out to the left. It looks like a hat that's upside down.*
- B. *The guy that's pointing to the left again?*
- A. *Yeah, pointing to the left, that's it! (laughs)*
- B. *Okay.*

Even though the first perspective proposed by A is understood by B, the conceptual pact that had been established between them reflected a different conceptualisation, which is ratified in its status by both interlocutors once it is brought back to the conversation. This dialogue fragment shows that the conceptual pact is one of a number of possible conceptualisations, while at the same time holding a special status as an agreed-upon tool that stands not as a direct description of the referent, but as a link to the dyad's past history of interaction.

Unlike linguistic precedents, conceptual pacts are not merely accumulated, but actively established by the interlocutors. In tangram games, alternative conceptualisations can be proposed by a participant only as a consequence of

having understood a previous conceptualisation by their partner, as they would reflect a different or complementary view on something that has already been identified. Past descriptions or references can be discarded or ratified only if they are understood as descriptions of the item.

A. *Okay, the next one looks, is the one with the person standing on one leg with the tail.*

B. *Okay.*

A. *Looks like an ice skater.*

B. *Yeah, okay.*

(Clark & Wilkes-Gibbs, 1986)

Considering this, we could assume that in referring to any object there should be a difference in provisionality (i.e. certainty with which the reference will be reused), derived from the level of grounding that can be assumed from the reference. A mutually confirmed conceptual pact would correspond to the highest level of grounding, with the mere use of a reference with no feedback that could confirm or disconfirm its understanding and/or acceptance, at the bottom of the scale. This difference would imply that, as recent studies have shown (Van Der Wege, 2009; Brown-Schmidt, 2012), a speaker with higher grounding certainty should have a stronger basis for interactional processes that rely on the certainty of the established convention, like the referential reduction process that interlocutors engage in after they ground a specific perspective (Clark & Wilkes-Gibbs, 1986; see also Chapter 2, Literature Review). However, it would also imply that a speaker with lower grounding certainty should be less inclined towards these interaction-dependant processes, and specifically, that the degree of grounding of the speaker's references should influence their reuse. Thus, we might ask how the addressee's participant role (their degree of involvement in the dialogue) might affect the speaker's certainty in her reference choices. This will be developed in the next section.

#### **5.1.4 Participant roles and co-presence heuristics**

An initial definition of who is and is not a participant in a conversation was proposed by Goffman (1974), differentiating an addressed participant from a ratified participant and from an overhearer. An addressee is defined as the individual to whom the speech is directed and who is expected to answer or give other forms of

feedback; while a ratified participant is still considered as part of the conversation but is not specifically addressed by the speaker. An overhearer, on the other hand, might or might not be acknowledged as present and/or encouraged to listen in. Research involving third parties in dialogue settings seems to vary in terminology, with some studies stating that silent individuals accessing the conversation are to be considered side-participants (Wilkes-Gibbs & Clark, 1992), and others describing any individual who is not a ratified addressee as an overhearer (Schober & Clark, 1989). From the point of view of the speaker, there is a social obligation towards the ratified participants in the conversation that is not present with respect to the non-ratified participants, overhearers or eavesdroppers (Clark, 1985). This social obligation entails that the utterances produced by the speaker must be designed with the intention of being comprehended not only by the addressee but also by all other ratified parties to the conversation, though not necessarily by outsiders.

According to Schober and Clark (1989), overhearers should be at a disadvantage, compared to ratified participants, even if they are members of the same cultural group as the interlocutors, the interlocutors do not know each other in advance, and they have witnessed the conversation from the beginning. This disadvantage would derive from the nature of the grounding process, which requires actions by both the speaker and the listener<sup>3</sup>. However, even if there is no guarantee of understanding, most references to common objects should be considered understandable, in the absence of confounds or ambiguities. Following this argument, Wilkes-Gibbs and Clark (1992) stated that co-presence and attention to what is being said should be enough for an individual to be considered a participant in the conversation. It is implied in this statement that references are considered to be transparent enough as not to require negotiation to establish their meaning; however, it is not clear how the speaker is supposed to interpret this participant's status.

Co-presence heuristics serve to short-circuit an important part of the workload that a complete recursive inference process would require if we were to track the full amount of common ground we share with our interlocutor in real-time conversation (Clark & Marshall, 1981; Barr, 2005). Indeed, if people share the same context and

---

<sup>3</sup> While this might be true for complex dialogue situations where the need for clarification or repair requests is generally high (see Healey, De Ruiter, & Mills, 2018), for simpler exchanges the difference between being the addressee for whom speech is tailored, and being an overhearer who has access to the same information but cannot actively interact with it, might not be as big.

speak the same language, in the absence of additional difficulties they should be entitled to believe they will understand what is established in the conversation in a similar way, *unless* what has been said is not sufficient to achieve shared mental representations (Brennan, 2005). Different *grounding criteria* are said to be established by speakers in each communication circumstance, depending on their evaluation of how much information should be shared as baseline between them, and other contextual considerations such as the communication medium and the perceived difficulty of the referential task.

Clark and Carlson (1982, see also Clark and Schaefer, 1989) proposed the *Principle of Responsibility* by which, after speakers have proposed a conceptualisation in the form of an utterance, the addressee needs to give positive evidence of its acceptance. After this process is completed, both interlocutors are assumed to be mutually responsible for their understanding of what was said. This agreement entails that the addressee is responsible for taking action when he has *not* understood the speaker, initiating a process of collaborative negotiation in order to reach understanding. The definition of positive evidence, in this context, includes all possible addressee actions that do not establish misunderstanding, such as continuing paying attention to the speaker, or any positive backchannels or feedback. A second *Principle of Distant Responsibility* was proposed by Clark & Wilkes-Gibbs (1986), to cover for those instances in which the speaker cannot obtain appropriate and timely feedback from their addressees, such as in writing, or video recording. According to this principle, speakers can self-repair or change their references according to their evaluation of their own output, without any input from the addressee.

Linguistic feedback in the form of either full turns by the addressee or backchannels are considered by the speaker as evidence of understanding, with different types of backchannels being consistently interpreted as indicating different communicative demands that need to be fulfilled, and thus ‘steering’ the course of the speech in different ways (Tolins & Fox-Tree, 2014). Considering this argument, consistent affirmative feedback (“*Yes, I know which one you mean*”, “*Got it*”, “*Done*”, or even shorter *recipency markers* like “*Yup*”, Bangerter & Clark, 2003) should be interpreted as signalling to the speaker that their descriptions are being understood as they are, and therefore are –all other things being equal– ‘good’ descriptions of

the target. For the speaker, this confirmation of the success of their utterances suggests that whatever strategy was employed in their planning and production was appropriate to both the addressee and the context where it was to be understood (Clark & Krych, 2004).

It could be considered that the original addressee's feedback might work as proof of 'general' comprehensibility for the speaker, who might extend their positive (or negative) evaluation of the reference in use to all other hearers who shared the same context (considering that a grounded perspective may be thought to be objectively easier to understand; Branigan et al., 2011). In this line, Gorman, Gegg-Harrison, Marsh, and Tanenhaus (2013) found that speakers used descriptions instead of names with a new addressee when informed that this addressee shared the same knowledge as a previous partner who was not aware of the object's name, suggesting speakers were able to extend their model of the addressee's knowledge to new interlocutors. Most studies so far (Brennan & Clark, 1996; Garrod et al., 2007; Van Der Wege, 2009) have shown speakers' descriptive choices are affected by their common history with their partner, such that they would be more likely to abandon previously used descriptions if they believe their current addressee does not share the history of use of that reference. In the following section, we discuss the design of two experiments, aimed at understanding the process by which speakers come to make referential decisions in the light of both their own production needs and their addressee's status in the conversation.

### **5.1.5 Experiment motivation and design**

How are the different factors that influence speakers' choice of referential expressions (production costs, lexical availability, provisionality of the expression, identity of the addressee, etc.) weighed in real-life dialogue? To specify this problem, we ask how the cost of producing an overspecified reference (in terms of cost for the speaker, and assuming no significant costs for the listener, Metzger & Brennan, 2003), with either the same or different partners, is weighed against the cost of switching to a new basic-level label, which should be easier to retrieve and produce (as discussed in section 5.1.2). Initially, we can consider four factors that should be relevant in this evaluation:

- 1) The strength of the precedent, that is, how many times it has been used and how strong is its association with the referent.
- 2) The degree of provisionality of the description, derived from the feedback and role of the conversational partner in the establishment of the precedent
- 3) The status of the conversational partner (new vs known partner) in the test phase (when the entrained descriptions are no longer justified by the context, as in Brennan & Clark, 1996, experiment 3)
- 4) The production cost of the descriptions, derived from the number of descriptions to be produced, ease of production (availability of salient dimensions, word retrieval, actual production/articulation), and cost of switching.

Brennan and Clark (1996, experiment 3) used a card sorting game with two stages to understand whether speakers maintained or changed previously entrained referential descriptions according to the identity of the interlocutor. In the first part of their experiment, participants sorted a set of cards where each of the target items was shown along with other items in the same category (i.e. a target dog with three other dogs in the set), requiring specific descriptions, and in the second part, they sorted a different set where the same target items were unique in their category (i.e. only one dog in the set). The participants varied in whether they would maintain their overspecific descriptions or not in this second part based on the identity of their interlocutor: if the interlocutor was the same as in the first part, participants kept the entrained terms even when this meant being over-informative in the current context (i.e. they kept on calling the dog card "*white terrier*" when "*dog*" was enough to identify the card in the new context), but they switched to basic-level terms if talking to a new interlocutor.

However, their study also showed that speakers across conditions kept using the entrained descriptions when these became overspecific in context, with over 70% of use of overspecific descriptions in the first round of the second part of the experiment, even for speakers with a new partner. Even in the fourth round of the second part, after having completed 3 rounds in the new context, 40% of

descriptions in the New Partner condition were overspecific. Why were Brennan and Clark's speakers not switching massively to more context-appropriate alternatives? The simplest explanation would be priming: participants' use of entrained descriptions was maintained because the lexical precedents they had established primed them to keep on using the same labels upon seeing the same referents, with relative independence from their conversational partner. However, there is a formulation cost associated to producing longer descriptions than needed, especially if these include non-basic-level labels, as these labels are assumed to be less readily available than unmarked basic-level labels, and might be also considerably longer (e.g. *"the small white terrier that's sitting"* vs *"the dog"*). How is this formulation cost weighed against the production cost of switching and using a new label?

To test this Production Cost assumption, we manipulated both the number of descriptions that needed to be produced, and the availability of salient dimensions that would be potentially easy to produce, particularly colour (as we discussed in the previous section, colour is the most likely dimension to be spontaneously included in speakers' descriptions). Overspecific descriptions should be more likely to be maintained if the salient colour dimension is available for description than if it is not (particularly when speaking to a new partner with whom there is no history of previous interaction), as colour is both easier to overproduce and not damaging for comprehension (Rubio-Fernández, 2016). Overspecific descriptions should be also more likely to be maintained if the number of references to be produced is low, as the potential overall gain from switching to basic-level descriptions would be higher if the number of descriptions to be produced is high. In this evaluation, speakers would need to consider, on one hand, the formulation costs of their lexical precedents (which should be lower if the overspecified dimensions are more easily available), and on the other, the potential gain of switching to and using basic-level labels (considering both the availability of 'unmarked' basic-level labels and a cognitive cost of abandoning primed descriptions<sup>4</sup>).

Brennan and Clark's results showed a difference in the use of referential precedents depending on whether the current addressee was present or not in the conversation

---

<sup>4</sup> It is assumed here that speakers would be self-priming overspecification across references, causing them to either maintain overspecific descriptions or to switch to basic-level terms across all target items, as reported on Goudbeek & Krahmer (2012).



where these precedents were established. That is, they compared the maximum possible grounding, where the speaker had complete confidence that the addressee shared this precedent with them, with no grounding at all. How would speakers behave if they had less certainty that their interlocutor had accepted this precedent? Considering the discussion in section 5.1.4, we suggest the speaker should maintain their entrained references to a higher degree if talking to the same partner as before, who had provided positive feedback of their understanding and acceptance of the reference; to a lesser degree if talking to a partner who had only witnessed their previous reference use, providing no feedback of their understanding; and to an even lesser degree if talking to a new partner, who had not witnessed their previous reference use nor been co-present with them.

To test this Provisionality assumption (the more certain the description's shared status, the more it will be kept), we designed a referential task similar to Brennan and Clark's, where a Director would describe a set of images to a Matcher, but including the figure of an Overhearer, who would have access to the conversation in the first half of the experiment without taking part in it and without being addressed in any way<sup>5</sup>. If provisionality affects the subsequent use of entrained descriptions, we should expect speakers to use less entrained descriptions when talking to the Overhearer in the second part of the experiment, compared to talking to the Same Partner they had in the first part. We should also expect speakers to use more entrained descriptions when talking to the Overhearer than when talking to a New Partner.

We designed two experiments to cover these hypotheses. Both experiments used confederate participants to fulfil the Matcher and Overhearer roles (see section 5.1.5.1). Experiment 4 presented participants with a small 12-item display of images, with 2 target images that appeared along with other images of the same category in the first three rounds of the experiment, and alone in their category in the second three rounds. 6 images in the set were colour photographs of objects,

---

<sup>5</sup> In practical terms, our Overhearers are somewhat in between what Wilkes-Gibbs and Clark (1992) defined as 'silent side-participant' and 'omniscient bystander': while they are not considered part of the conversation nor addressed in any way, they can see all the Director's movements and utterances through a duplicated-screen monitor, and they are seated in a separated position at the same table as the Director. In Wilkes-Gibbs and Clark's study, results from participants interacting with partners who had been in either of these two conditions were significantly different from results from participants with new partners.

and the other 6 were black tangram figures. Participants played the 6 rounds of the game as Directors, with confederates taking the roles of Matcher and Overhearer. They were located in one of three conditions: A Same Partner condition, where they played the second half of the game with the same partner they had in the first half; an Overhearer condition, where they played the second half with the overhearer they had on the first half; and a New Partner condition, where they played the second half of the experiment with a new confederate Matcher.

Experiment 5 presented participants with a larger 20-item display, with 4 target images that appeared along other members of the same category in the first half of the experiment, and alone in their category in the second half. 12 of these images were monochrome drawings of objects, and the other 8 were black tangram figures. Experimental conditions were the same as in Experiment 4.

The structure of the experiments implies that there are two crucial points of measure: the first one is the difference between round 3 (the last round of the first half) and round 4 (the first round of the second half), as this is the point in which participants are faced with a new context, and with either the same or a different partner. The second measure is the difference between the rounds in the second half (4 to 6), as participants might choose a different reference after they have had time to process the new conditions (as in Brennan and Clark, 1996).

As per our Production Cost assumption, we should expect participants to use more overspecific descriptions in Experiment 4, where the colour dimension is available and the number of items to be described is low, than in Experiment 5, where no dimension is salient in overall terms and the number of items to be described is higher.

As per our Provisionality assumption, we should see a gradient of maintenance of overspecific descriptions in our Participant Role conditions, with participants using more overspecific descriptions in the Same Partner condition, less in the Overhearer condition, and significantly less in the New Partner condition. Moreover, we should see less switching to basic-level descriptions throughout the second half of the task (rounds 4 to 6) in the Same Partner condition, compared to the other two conditions, as the certainty of the 'shared status' of their initial descriptions should be highest in

this condition, and this evaluation should not change throughout the task. Both the Overhearer and the New Partner condition should show more basic-level descriptions throughout the second half of the task.

#### **5.1.5.1 Use of confederates**

Our experimental setup uses confederates in the roles of both Matcher and Overhearer; the Director was always a naïve participant. Even though some researchers have expressed concern over the use of confederates in dialogue conditions (Kuhlen & Brennan, 2013), our setup prevented the most common problems that their inclusion might cause, i.e. inadvertently shaping participants' behaviour through verbal or nonverbal cues such as facial expressions, body posture, tone of voice, pauses, etc.; and participants guessing their confederate status. In our experiments, confederates were instructed to respond as naturally as possible to their interlocutor's output, and (in the Matcher role) to avoid including any conceptually relevant information that had not been included previously by the Director when asking for clarification, opting instead for more general questions (such as "*I don't really get which one you mean*", or "*Could you be a bit more specific?*"). Moreover, the fact that almost all their verbal interaction was performed through a written chat reduced the possibility of non-verbal feedback that could escape the confederate's conscious control. The specifications of the Overhearer role, where non-verbal feedback could have occurred, prevented this from happening, as the Overhearer was seated sideways from the Director's position (thereby not in direct view), and instructed not to speak or communicate with the Director at all, unless addressed. In the few instances where the participant addressed the Overhearer (most of the time to share a joke, or to comment on the difficulty of the task), confederates were instructed to respond naturally and to say no more than what would be minimally necessary to complete the interaction in a natural way.

## **5.2 Experiment 4**

### **5.2.1 Participants**

We tested 30 students from the University of Edinburgh, 23 females and 7 males. Data from one participant was discarded as they did not follow the instructions (they supplied overly ornate descriptions that avoided mentioning what the item actually was). Participants were paid £5 for their time (usually between 25 and 35 minutes). Confederates were asked to inform the experimenter if they knew any of the participants; this did not occur. 10 participants were assigned to each experimental condition.

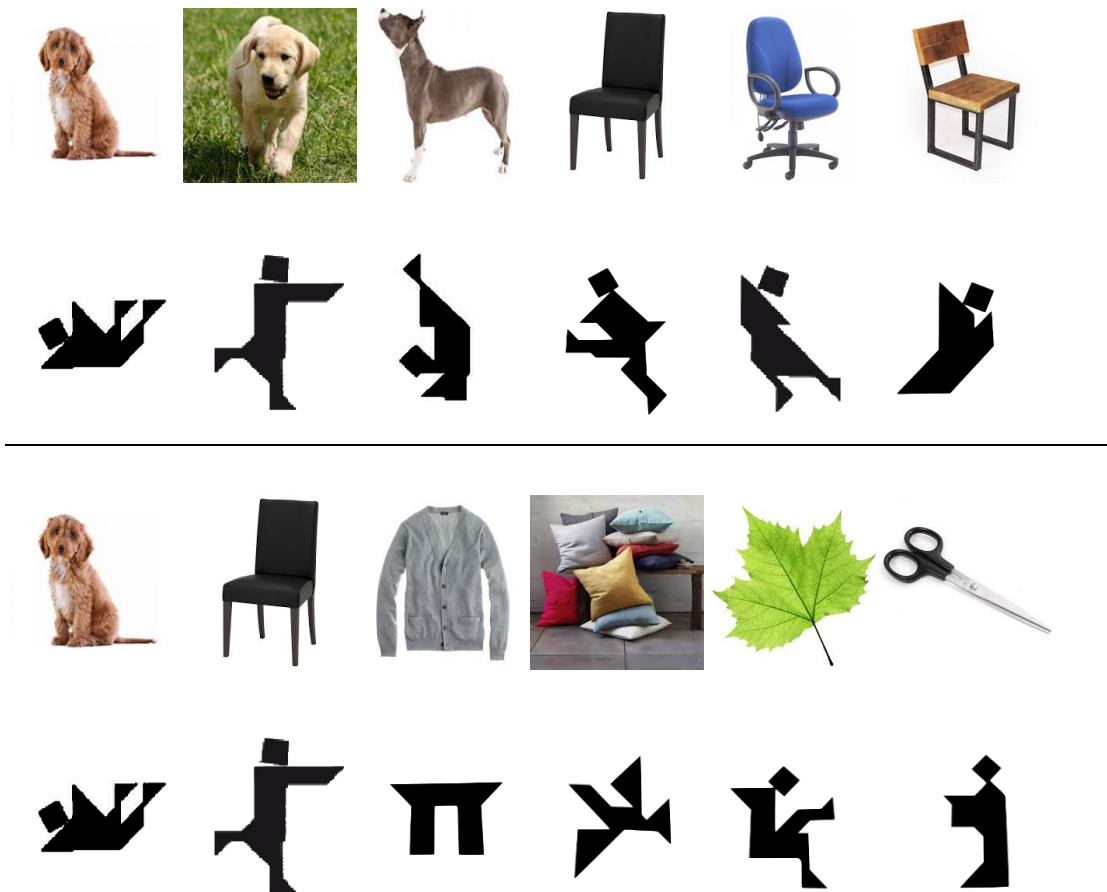
Three female voluntary Research Assistants acted as confederates, with two of them playing the roles of Matcher and Overhearer alternately with each new participant, and a third playing the role of new Matcher in the New Partner condition. They were trained in how to respond to participants as explained in section 5.1.5.1, but did not receive any further instructions during the experiment, where they were treated as naïve participants. They were not informed of the hypotheses of the experiment until after the testing was completed.

### **5.2.2 Materials**

The experiment setup involved two sets of images, A and B, with 6 colour photographs of common objects and 6 monochrome tangram figures per set (Fig. 5.1). In each A set, two common object images were targets, and these appeared as well on the corresponding set B. All object images in each A set belonged to two categories, with one target item and two fillers per category (e.g. one target dog and two other dog images). The target item was selected as being the closest example of a prototypical image for that category. The object images in set B included the two target items from set A, and 4 non-related images of common objects (a cushion, a cutlery set, a bird, etc.). The tangram images, included to distract participants from the changing figure/ground relationship of the object images in the two sets (as suggested in Brennan & Clark, 1996), were similarly distributed. Set B

included two tangrams from set A, and 4 non-related tangram figures that were new to participants. This distribution was similarly aimed at distracting participants from the change in context, promoting the alternative idea of a partial change across the board, where some images were maintained and some were changed.

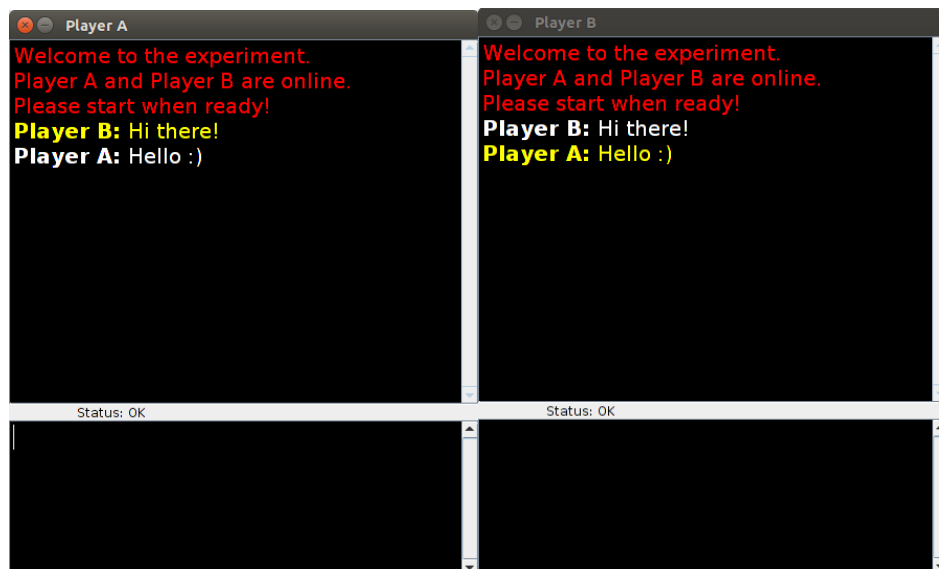
We designed 6 pairs of sets: 2 sets of *dog* and *chair* categories, 2 sets of *leaf* and *shoe* categories, and 2 sets of *fish* and *mug* categories. Set A was used in the first 3 rounds of the experiment, and set B was used in the second 3 rounds. Set pairs were used equally frequently in all conditions, in random order.



**Fig. 5.1** Set A example (top) and Set B example (bottom).

Images were randomised in each round and displayed in two rows of 6 200x200 px. images each. A separate window showed a chat tool, similar to the ones on

commercial software (Facebook, Whatsapp). The chat window (Fig. 5.2) displayed a text at the beginning of each round which informed the participant of an image set being loaded, included to establish a specific time delay between rounds (2 seconds). At the end of round 3, the participant was additionally informed that he/she would be playing the remainder of the experiment with either a) the Same Partner they had on the first part, b) the Overhearer they had on the first part, or c) a New Partner. This message included a longer delay of 6 seconds, to allow participants to notice the change in the image set. Participants could chat freely with their partner (however Matchers were instructed to keep their utterances brief and to avoid adding new conceptualisations to the conversation, as described in section 5.1.5.1).



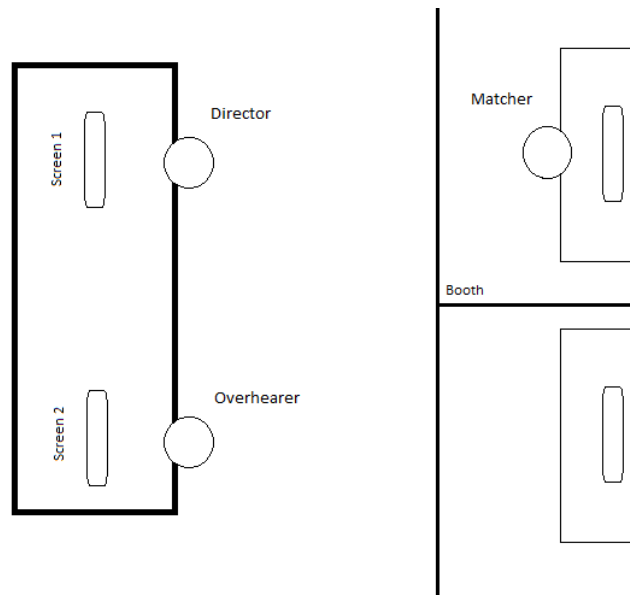
**Fig. 5.2** Chat displays for Director (Player A) and Matcher (Player B).

### 5.2.3 Design

The experiment had a 3x2 design, with factors Condition (Same Partner vs Overhearer vs New Partner, between participants) and Block (rounds 1-2-3 and 4-5-6, within participants). However, the analyses were performed on either Round 4 alone (compared to round 3), or Rounds 4 to 6.

#### 5.2.4 Procedure

Participants came to the lab individually. Confederates entered the lab from an adjacent room 5 minutes before the scheduled time; if the participant came in after that, they were informed the confederates were the other participants that got to the lab a few minutes earlier, if the participant came in before that, the confederates left the adjacent room through an exit door and came to the lab through the main door, either individually or casually together. As soon as both the participant and the confederates were there, they were instructed to take their positions in the experimental setup; they were all treated like naïve participants throughout the experiment. The participant was seated in the Director's position on the right side of a large desk with a network computer and a flat screen, while the Overhearer was seated on the left side of the same desk, with an additional screen that mirrored the display of the participant's screen (Fig. 5.3). They were both informed that their screens were mirrored, meaning the Overhearer could see the Director's images and chat (though the images were not numbered or highlighted in any way as the Director described them, so their identification rested on being able to interpret the descriptions correctly), but that the Overhearer was not to participate in the conversation nor to interact in any way with either the participant or the Matcher during the first part. Specifically, the participant was told the Overhearer would be observing their performance in the task so they could either perform the same task at a later point (in both the Same Partner and the New Partner conditions), or perform as Matchers in the second part of the task (in the Overhearer condition). The confederate Matcher was seated in a booth, with a network computer and a flat screen. In the New Partner condition, an additional confederate was instructed to wait for their time to participate in another booth.



**Fig. 5.3** Task set-up. Top left: Director position. Top right: Matcher position. Bottom left: Overhearer position. The Overhearer had access to a screen that mirrored the display of the Director's screen, but could not see the Director's screen directly.

Both the participant and the confederates were then handed a written copy of the instructions and asked to read it. The experimenter then took a moment to briefly summarise the procedure of the experiment, and to answer any questions participants might have had. At this point, participants were informed that they were going to play the second part of the experiment with either the same partner as on the first part (Same Partner condition), the overhearer they had on the first part (Overhearer condition), or a New Partner (New Partner condition). They were told the Director was to describe the set of images to the Matcher in order, in a simple and straightforward way, and that the Matcher was to place the corresponding images in the appropriate positions. They were also informed they could interact freely using the chat. After this, they were instructed to begin playing, and the experimenter left the room and went to the adjacent room.

The Director's display showed a 2x6 grid with the 12 images of a randomly chosen set A, in a quasi-random order that prevented the target images from being placed first or last in the array. The same display was showed through mirroring in the Overhearer's screen. The Matcher's display showed 2 2x6 grids, an empty grid on the top and a full grid on the bottom, with the 12 images of the same set A in a



random distribution. Both displays included a chat panel, divided into a bottom writing space and a top dialogue space, showing up to 4 lines of the preceding chat dialogue (see section 5.2.2). This chat panel showed both the dialogue between the participants, and the messages from the server, in different colours.

### 5.2.5 Coding

Brennan and Clark (1996) defined three aspects that were crucial in their analysis of participants' use of conceptual pacts: whether participants used the *same* description as in the previous round (a measure of lexical entrainment); whether they used *basic* descriptions or not (anything but a basic description would have been overspecific in the second half of the task); and whether their descriptions were reduced in length compared to the same reference in the previous round.

We analysed and coded all target item responses for 29 participants in the 6 rounds of the experiment, considering three dichotomous measures:

1) Same: Measure of lexical entrainment; whether the description of an item in a given round is the same as the description of the same item in the previous round. We defined 'sameness' as Brennan and Clark (1996), considering all content words in a description, and not considering hedges or modifiers, nor order, i.e. *the fluffy dog that's standing* would be considered the same as *the standing fluffy dog*, but not the same as *the fluffy dog*.

2) Basic: Measure of use of basic-level (*unmarked*, on Cruse's (1977) terms) labels; whether the description of an item is the basic-level label associated with that item, i.e. dog, car, fish, etc. Basic-level descriptions should not be used in the first part of the experiment, as that would make identification impossible.

3) Reduced: Measure of description length reduction; whether the description of an item is reduced in length (number of words, considering all content words) with respect to the description of the same item in the previous round. If the description was lexically different, it was nevertheless compared in number of words to the previous description for the same item. If a description of length=1 was followed by

another description of length=1, this second description (and any further length=1 descriptions that followed) was excluded from the analysis, as by definition they can't be reduced. Descriptions that were augmented (which occurred rarely, and mainly in Round 4) were excluded from the analysis.

Considering the task procedure, the main measuring points are the first round of the second half, Round 4 (the point at which participants acknowledge both the new context and the fact they are interacting with either the same or a different partner), and the trajectory throughout the rounds of the second half, Rounds 4 to 6.

### **5.2.6 Results**

Considering the critical measure points described in the previous section, we selected our models to account for both the main point of change, Round 4, and the trajectory of change across the second half of the task, Rounds 4 to 6. In round 4, the change from a more complex to a simpler context where each target image was unique in their category rendered the previous references overspecific, as in this new context basic-level labels should have been sufficient to uniquely identify each referent (e.g. whereas in round 3 *"the dog"* would be insufficient to uniquely identify the target dog from the other dogs in the set, on round 4 such a description would be sufficient). Moreover, this was the point at which our experimental conditions were differentiated, with participants moving on to play with either the Same Partner as before, the Overhearer, who had witnessed their previous interaction, or a New Partner. On the other hand, the process of switching to a basic label might take place over more than one turn (e.g., reducing a label from *"the fluffy white dog"* to *"the white dog"* in round 4, but reducing it further to its basic form *"the dog"* in round 5), as participants might need the experience of acknowledging the discrepancy between the level of reference that the context requires (in terms of amount of information provided), and the level of reference they are producing (as was the case in Brennan and Clark, 1996).

We selected three Round 4 models, testing for the use of the Same description (compared to Round 3), Basic-level descriptions, and Reduced descriptions (compared to Round 3); comparing the output of our participants according to their

Condition (Same Partner vs Overhearer vs New Partner). We used logistic multilevel models (lme4 package in R, Bates et al., 2015) to evaluate the effect of Condition over our measured variables. For rounds 4 to 6 models, Condition was deviation-coded. We initially included Participant, Image, and Image set, as random effects, but both Image and Image set proved to be close to null across models, and were therefore eliminated from the model structure. All p-values reported come from Likelihood Ratio Tests performed by fitting the full models using the *mixed* function in the afex package (Singmann et al., 2017).

We also tested for effects across Rounds 4 to 6. For these models, we included both Condition and Round number, coded as a numeric predictor, plus the interaction between the two. Random effects structure was similarly reduced to a by-Participant intercept. For each variable, I will present the results of the Round 4 models first, followed by the results of the models for Rounds 4 to 6.

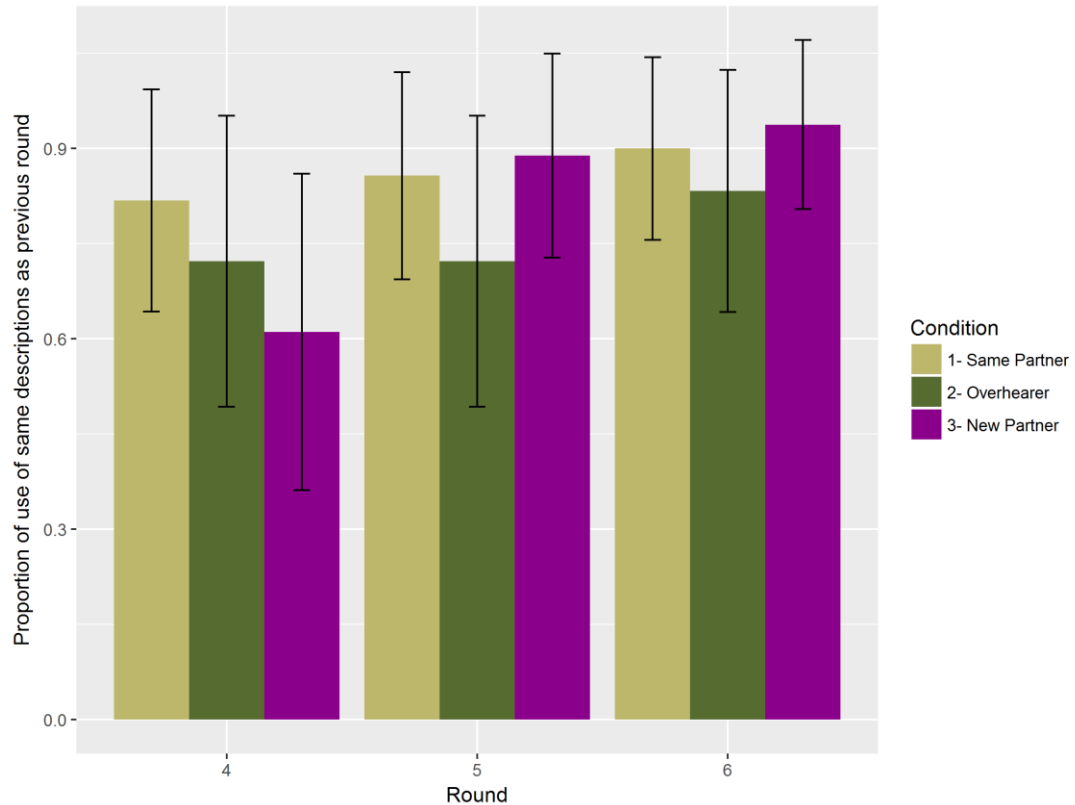
*Use of Same description - Lexical entrainment.* Participants used a high proportion of previously used references in Round 4, in all conditions: 81.8% in Same Partner condition, 72.2% in Overhearer condition, and 61.1% in New Partner condition (see Fig. 5.4).

Condition was not a significant predictor of the use of the same references in Round 4: even though there was a numeric difference showing more use of entrained descriptions in the Same Partner condition and less in the New Partner condition, this difference was not significant ( $\chi^2=2.13$ ,  $p=0.34$ ).

In Rounds 4 to 6, participants used increasingly more entrained descriptions as they advanced in the game, again with no significant differences between conditions. Round was a significant predictor of the use of the Same vs a Different description ( $\chi^2=6.00$ ,  $p=0.01$ ), across Conditions. Neither Condition ( $\chi^2=2.43$ ,  $p=0.30$ ) nor the interaction of Condition and Round ( $\chi^2=2.50$ ,  $p=0.29$ ) were statistically significant.

Considering previous research (Brennan & Clark, 1996) and our own experimental predictions pointing to any differences between the Same Partner and the New Partner conditions being larger than the difference between any of those two conditions and the Overhearer condition, we ran a second model including only the

Same Partner and the New Partner conditions. This model showed a significant increase in the use of the Same descriptions with Round ( $\chi^2=5.70$ ,  $p=0.02$ ), across Conditions, but again no main effect of Condition ( $\chi^2=2.09$ ,  $p=0.15$ ), and no significant interaction between the two ( $\chi^2=1.77$ ,  $p=0.18$ ).

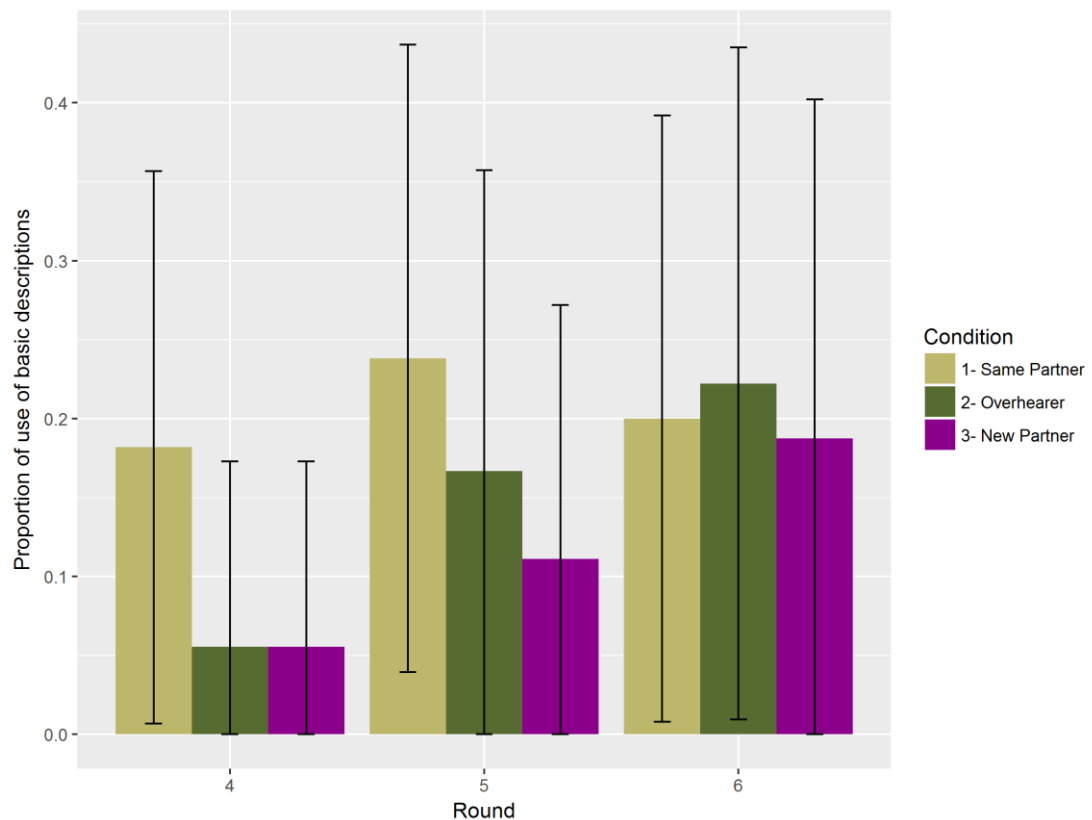


**Fig. 5.4** Proportion of use of the same description as in the previous round, Rounds 4 to 6, per Condition. Error bars show 95% CI.

*Basic level terms.* The use of basic-level terms in Round 4 was low in all conditions, with no significant differences between Conditions ( $\chi^2=2.27$ ,  $p=0.32$ ). Participants in the Same Partner condition used 18% basic-level descriptions in Round 4, with 5.5% for both Overhearer and New Partner conditions (see Fig. 5.5).

Considering rounds 4, 5, and 6, Round had a marginally significant effect over the use of more basic level terms, across Conditions ( $\chi^2=4.01$ ,  $p=0.05$ ): Participants were marginally more likely to use basic-level descriptions as they advanced in the game. There was no significant difference between Conditions ( $\chi^2=1.83$ ,  $p=0.40$ );

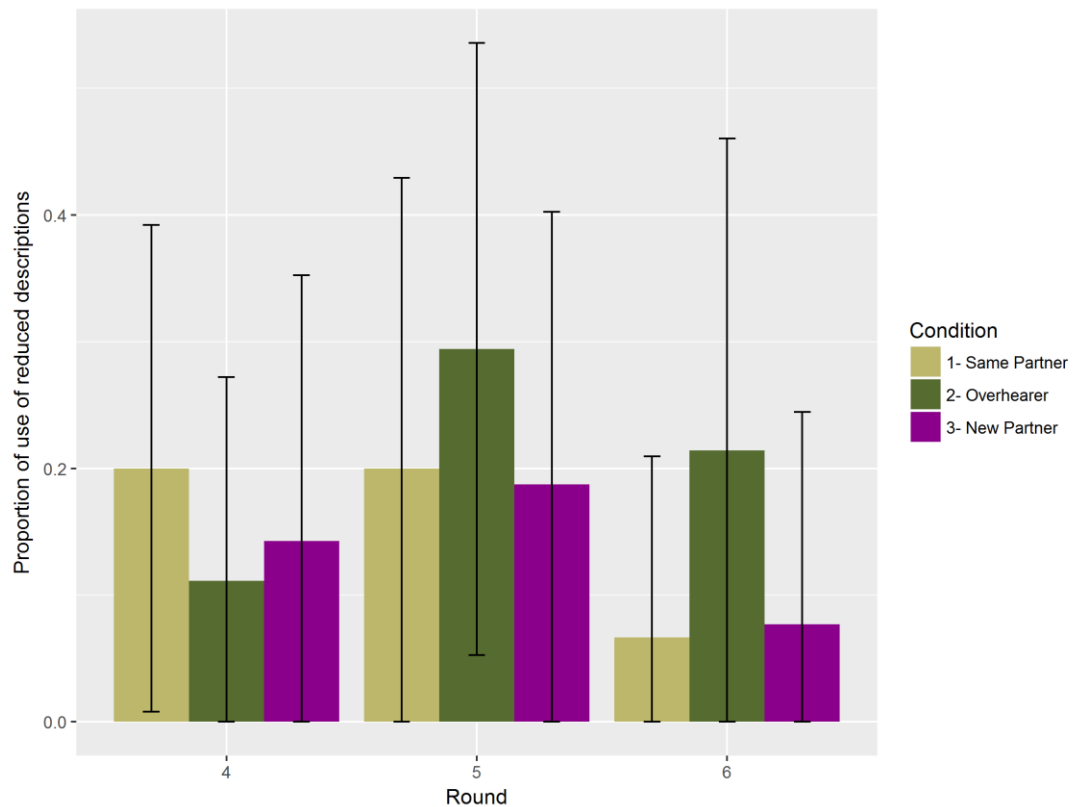
and the interaction between Round and Condition was similarly not significant ( $\chi^2=1.29, p=0.52$ ).



**Fig. 5.5** Proportion of use of basic-level vs overspecific descriptions in rounds 4 to 6 (second part), per Condition. Error bars show 95% CI.

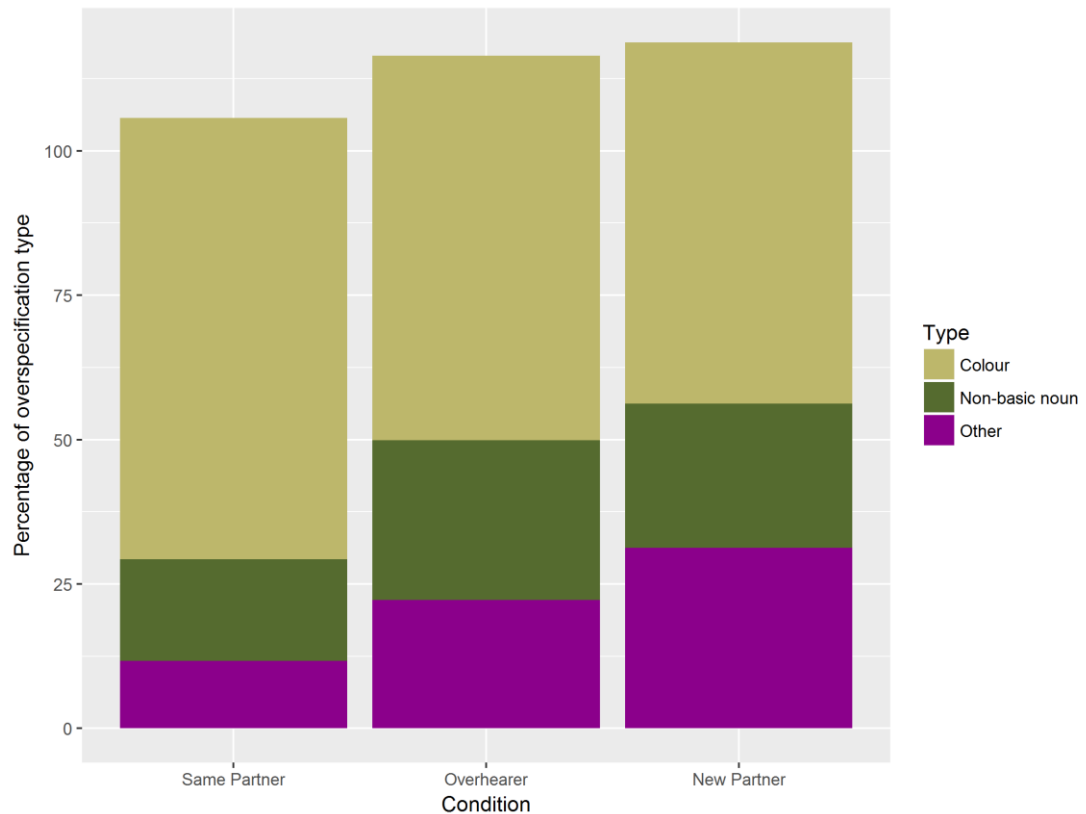
*Reduced description length.* The use of reduced-length descriptions in Round 4 was low, with no more than 30% of reduced descriptions in any condition. There were no significant differences in reduction of description length between Conditions ( $\chi^2=0.80, p=0.67$ ) (see Fig. 5.6).

Considering rounds 4 to 6, neither Condition ( $\chi^2=0.58, p=0.75$ ) nor Round ( $\chi^2=0.11, p=0.74$ ), nor their interaction ( $\chi^2=1.85, p=0.40$ ) were significant predictors of the use of more reduced descriptions: Participants maintained the length of their descriptions throughout the rounds, and there were no significant differences across conditions.



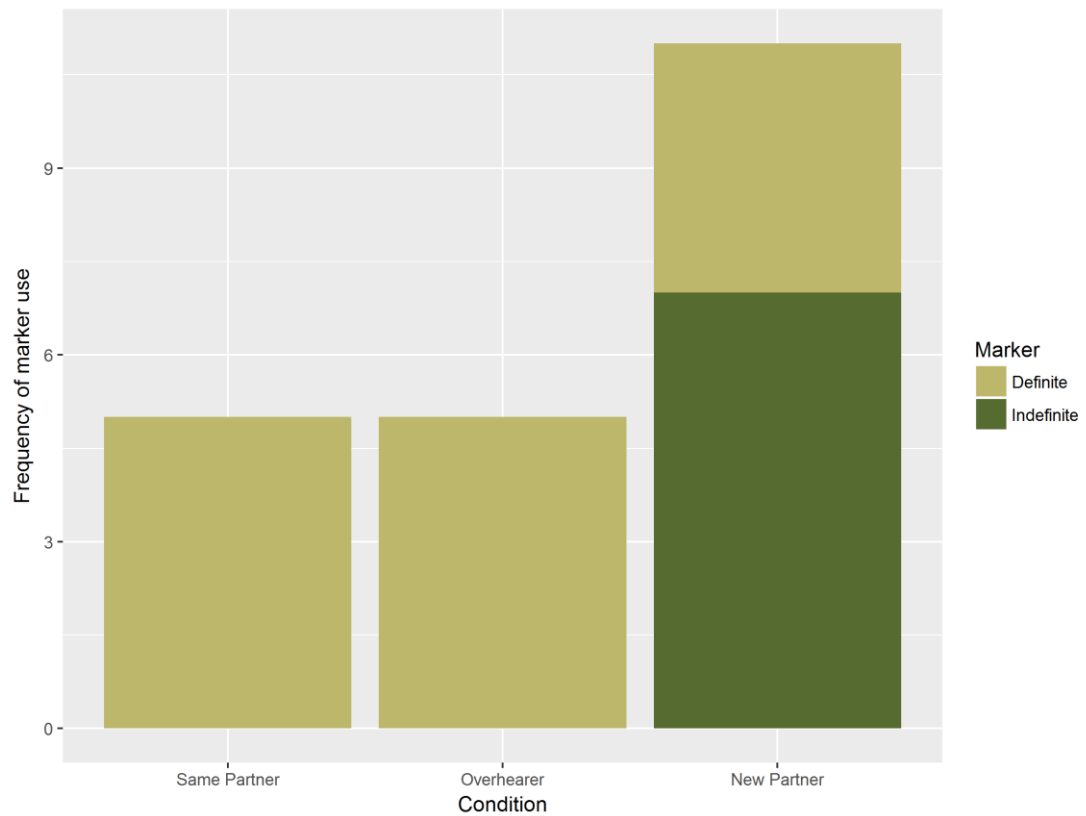
**Fig. 5.6** Proportion of use of reduced descriptions per rounds 4 to 6, per Condition. Descriptions of length 1 that were preceded by another length=1 description were not considered, as they could not possibly be reduced; therefore the proportion of Maintained descriptions only considers those in which length>1.

Additionally, we analysed overspecific target responses in round 4 to identify the type of overspecification they included, using three categories: Colour attribute, Overspecific (non-basic) noun, or Other (Fig. 5.7). In the Same Partner condition (10 participants, 17 overspecified responses), 76.4% of overspecific responses included a Colour attribute, 17.6% an Overspecific noun, and 11.7% Other attributes. In the Overhearer condition (9 participants, 18 overspecified responses), 66.6% of overspecific responses included a Colour attribute, 27.7% an Overspecific noun, and 22.2% Other attributes. In the New Partner condition (10 participants, 16 overspecified responses), 62.5% of overspecific responses included a Colour attribute, 25% an Overspecific noun, and 31.25% other attributes. Overall, 68.6% of overspecific responses included a Colour attribute, 23.5% an Overspecific noun, and 21.5% Other attributes (some responses included more than one overspecification category).



**Fig. 5.7** Percentages of overspecification type in Round 4, per Condition. Some responses included more than one overspecification category (e.g. “brown flat shoe”), therefore sum of percentages might be over 100.

We also analysed the use of definite vs indefinite markers in round 4, to establish whether participants switched their mode of reference, even if they didn’t change their lexical choices. In the Same Partner condition, 29.4% of all descriptions included a definite marker (“the”), and no descriptions included indefinite markers (“a”) (raw frequencies in Fig. 5.8). In the Overhearer condition, 27.7% of all descriptions included definite markers, and no descriptions included indefinite markers. In the New Partner condition, 25% of all descriptions included definite markers, and 43.7% of descriptions included indefinite markers. The use of definite vs indefinite markers was significantly different between Conditions (proportional test,  $\chi^2=9.54$ ,  $p=0.008$ ).



**Fig. 5.8** Frequency of marker use in Round 4, per Condition.

### 5.2.7 Discussion

Results show participants across Participant Role conditions maintained their entrained descriptions throughout the second part of the experiment, where the context made these descriptions overspecific. They did so even in the New Partner condition, where the descriptions had not been used before in the context of that partner, suggesting an intrinsic motivation to maintain their use (Van Der Wege, 2009). Round number was a marginally significant predictor of the use of more basic-level descriptions, with no significant differences between conditions, showing a slight increase in the use of context-appropriate descriptions over time. The use of entrained descriptions (same description as in the previous round) also increased with Round, showing participants across conditions tended to retain the description they had used previously. Taken together, results suggest most participants maintained the overspecified descriptions they were using throughout the second half of the experiment, with a reduced number of participants switching towards basic-level labels primarily in Round 4.



A more detailed analysis of the overspecific descriptions produced by participants revealed that colour was the most frequently overspecified dimension, with descriptions of the type *colour adjective + noun* being most frequent (more than 60% of overspecified descriptions included colour adjectives, across conditions). Previous studies have found that colour is the aspectual dimension most likely to be overspecified in descriptive statements, likely due to its perceptual salience (Tarenskeen et al., 2015, Rubio-Fernández, 2016). Moreover, colour has been suggested to be intrinsically easier to describe than other object dimensions such as size, requiring low cognitive effort to be perceived (Pechmann, 1989; Viethen, Goudbeek, & Krahmer, 2012).

Even though it cannot be proved that overspecification facilitated comprehension in this context, we can assume overspecification didn't impair communication either (as in Gann & Barr, 2014), as most of the overspecific descriptions involved colour adjectives, which has been suggested as potentially improving identifiability of referents (Rubio-Fernández, 2016). In this sense, overspecification should not be communicatively costly for the speaker nor threaten communicative success. Indeed, Grodner and Sedivy (2011) showed that listeners rapidly adapted their expectations when interacting with a speaker that frequently overspecified their references without a contrastive intent, suggesting listeners are not troubled by overspecification, especially if it is speaker-consistent. Our results suggest speakers are highly sensitive to speaker-internal pressures, especially when this does not interfere with successful communication (Wardlow Lane & Ferreira, 2008). In this sense, the use of indefinite markers in Round 4 in the New Partner condition, absent in the other conditions, confirms that speakers were aware they were interacting with new speakers who didn't share their precedents, but chose to use them anyway.

Our results show a different pattern from Brennan and Clark's (1996) study, who showed that a proportion of participants abandoned their overspecific descriptions over the course of the second part of the experiment, favouring basic-level labels. However, as their setup allowed free oral communication between participants, it is possible that feedback from the Matchers played a role in the Directors' decision to switch to simpler labels (as their own discussion suggests). In our experiment,

feedback from (confederate) Matchers was limited both as the written medium prevented them from using anything other than full turns (eliminating the possibility of backchannels and other forms of feedback), and per design, as confederate Matchers were instructed not to comment unnecessarily on Directors' responses.

Another important difference between our experimental setup and Brennan and Clark's is the number of items that needed to be described. A reduced set of items, as in our setup, implied that in the second part of the experiment only 6 items and 6 tangram figures needed to be described in each round, lowering the overall processing difficulty of the task. In a similar 5-item display experiment, Gann and Barr (2014) found participants were likely to overspecify their descriptions independently of the identity of their interlocutor (same partner vs new partner), and without considering the availability of feedback. Even though their findings point to an impaired comprehension effect for overspecific descriptions, this effect might be related to the fact that they used monochrome stimuli, where the additional information given in the overspecified descriptions had to do with properties of the items that are less salient in context than colour, like the 'meltedness' of a candle (melted/unmelted), the type of a guitar (electric/acoustic) or the material of a garbage bin (metal/plastic). In our experiment, there should have been no 'referential penalty' or extra processing cost (Wardlow Lane & Ferreira, 2008) associated to using overspecific references, especially considering the information most likely to be included in the overspecification –colour– could potentially help in the correct identification of the referents.

Our results suggest that, when the reconceptualisation cost of switching, added to the potential gain of using basic-level labels, is perceived as higher than the formulation cost of maintaining overspecific references, speakers are not likely to abandon their entrained descriptions. If the number of items to be described is low, the formulation effort implied in producing previously-used overspecific labels can be perceived as lower than the cost of reconceptualisation that is implied in switching to basic-level labels. Moreover, in our experiment these overspecific descriptions were mostly composed as colour adjective + noun, which, considering the ease of production of colour adjectives, could imply that the extra formulation effort involved in their production was not much higher than the cost of formulation of bare nouns.

Under what conditions would speakers switch to basic references? The review of the literature suggested two factors that seemed to influence this decision: the ease of production of overspecific references, where availability of salient dimensions would lower formulation effort, and the overall production cost of the set of descriptions, where a lower number of descriptions to be produced would imply lower effort. We manipulated these two dimensions in the next experiment, by eliminating the most salient dimension for overspecification (colour), and increasing the number of items to be described.

### **5.3 Experiment 5**

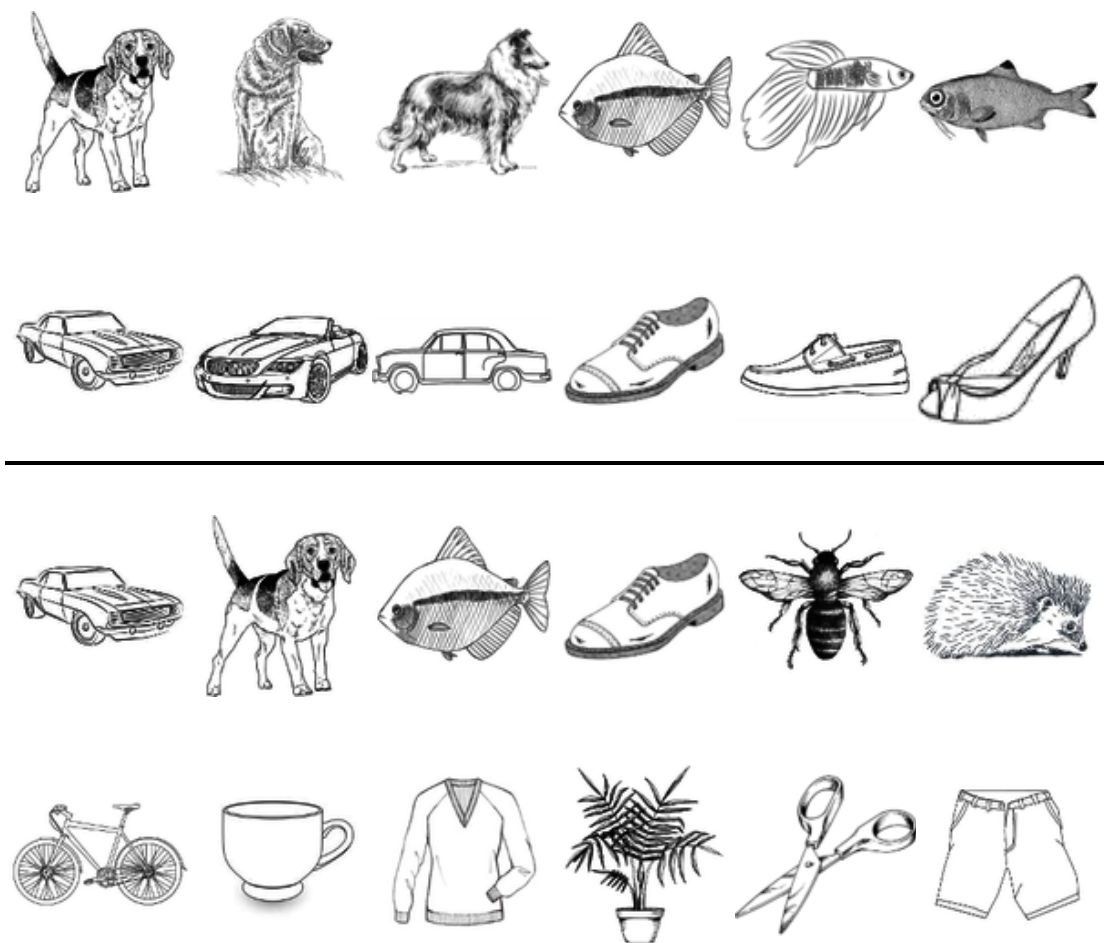
#### **5.3.1 Participants**

We tested 31 students from the University of Edinburgh, 17 females and 14 males. Data from two participants was discarded as they did not follow the instructions (one kept on asking unrelated questions to the Matcher and providing intentionally 'funny' descriptions, the other didn't follow the description order and only described target items in one of the three first part rounds, referring to them as "the last x" the rest of the time). Participants were paid £7 for their time (usually between 30 and 45 minutes). Confederates were asked to inform the experimenter if they knew any of the participants; this did not occur. 10 participants were assigned to the Same Partner condition, 11 participants to the Overhearer condition, and 10 participants to the New Partner condition.

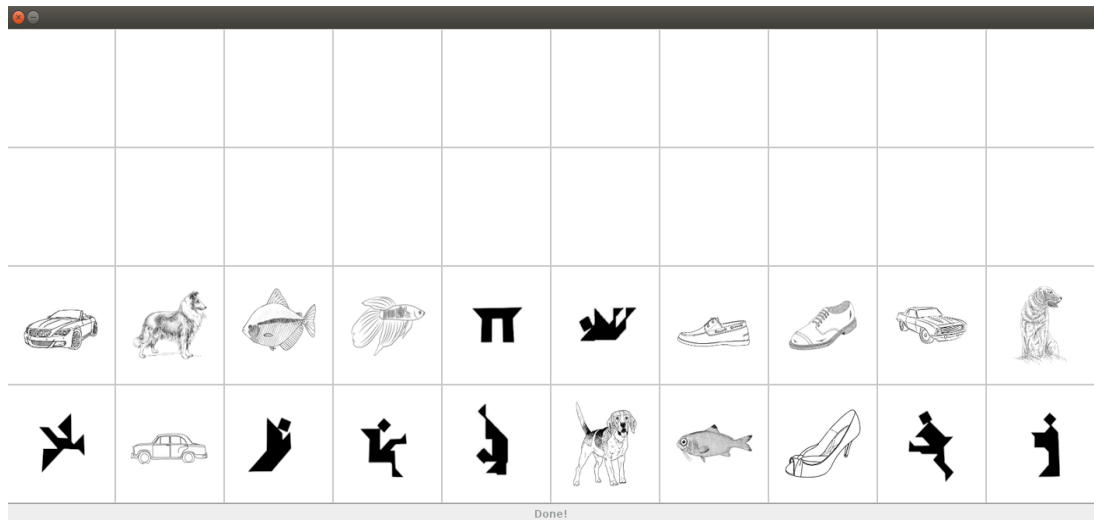
#### **5.3.2 Materials**

As in Experiment 4, the experiment setup involved two image sets, A and B, this time with 12 monochrome drawings of common objects and 8 monochrome tangram figures per set (Fig. 5.9). We designed 2 pairs of sets, comprising four categories of objects each (dog, car, fish, shoe). A set A was used in the first 3 rounds of the experiment, and the corresponding set B was used in the second 3 rounds. Set pairs were used equally frequently in all conditions, in random order.

All object images in set A belonged to the four categories, with one target item and two fillers per category (e.g. one target dog and two other dog images) (Fig. 5.10). The target item was selected as being the closest example of a prototypical image for that category. The object images in set B included the two target items from set A, and 8 non-related images of common objects (a jumper, a bicycle, a bee, etc.). Eight tangram images were included in each set, as in Experiment 4. Set B included four tangrams from set A, and 4 non-related tangram figures that were new to participants.



**Fig. 5.9** Set A example (top) and Set B example (bottom). Tangrams (not shown) were similar to the ones in Experiment 4.



**Fig. 5.10** Example of image set A, Matcher's view. Director's view includes only the two bottom (occupied) rows, in a different order.

Images were randomised in each round and displayed in two rows of 10 120x120 px. images each. All other materials were presented in exactly the same way as in Experiment 4.

### 5.3.3 Design

As in Experiment 4, the analyses were performed on either Round 4 alone, or Rounds 4 to 6.

### 5.3.4 Procedure

The procedure was identical to Experiment 4, with participants coming to the lab individually and interacting with a confederate Matcher in the first part, and either the same Matcher, a confederate Overhearer (who had witnessed the first part of the experiment in an additional mirrored screen), or a confederate New Matcher in the second part of the experiment.

### 5.3.5 Coding

The coding procedure was identical to Experiment 4, with measures for Same vs Different descriptions (lexical entrainment), Basic vs Overspecific descriptions (switch to context-appropriate descriptions), and Reduced vs Maintained descriptions.

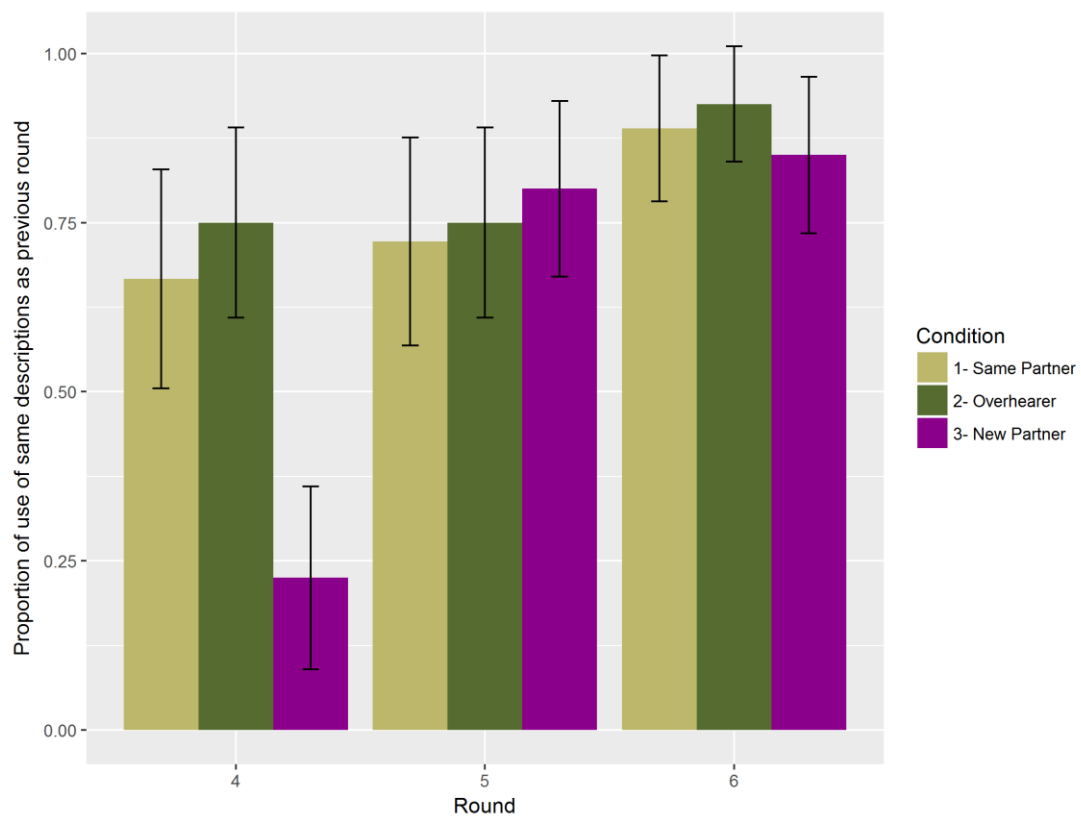
### 5.3.6 Results

As in Experiment 4, we used logistic multilevel models (lme4 package in R, Bates et al., 2015) to evaluate the effect of Condition and/or Round over our measured variables. Condition was dummy coded, with Same Partner as base level. We initially included Participant, Image, and Image set, as random effects, but both Image and Image set proved to be close to null across models, and were therefore eliminated from the model structure. All p-values reported come from Likelihood Ratio Tests performed by fitting the full models using the *mixed* function in the *afex* package (Singmann et al., 2017); the function was used once to obtain the significance values of interactions and main effects, and again with a parameter specification to obtain the significance values of each level of the factors in the interaction.

*Use of Same description – Lexical entrainment.* Condition significantly predicted the use of the Same terms (main effect of Condition  $\chi^2=14.32$ ,  $p<0.001$ ): Participants used the same terms in Round 4 (compared to Round 3) significantly more in the Same Partner condition than in the New Partner condition ( $\chi^2=8.65$ ,  $p=0.003$ ). The difference between the Same Partner and the Overhearer conditions was not significant ( $\chi^2=0.44$ ,  $p=0.51$ ).

For Rounds 4 to 6, there was a significant interaction between Round and Condition ( $\chi^2=7.90$ ,  $p=0.02$ ). Moreover, Round had a significant main effect over the use of the Same or Different descriptions, with participants using more entrained descriptions with each Round played ( $\chi^2=5.0$ ,  $p=0.03$ ), across Conditions. Condition as a main effect was also significant ( $\chi^2=17.48$ ,  $p<0.001$ ).

As the plot shows (Fig. 5.11), participants in the New Partner condition diverged from referential precedents in Round 4, and then maintained these new references in Rounds 5 and 6: The interaction was significant for the New Partner condition ( $\chi^2=5.50$ ,  $p=0.02$ ), but not for the Overhearer condition ( $\chi^2=0.01$ ,  $p=0.91$ ), compared to the Same Partner condition. The New Partner condition also differed significantly from the Same Partner condition across Rounds ( $\chi^2=10.36$ ,  $p=0.001$ ), while the difference between the Overhearer and the Same Partner conditions was not significant ( $\chi^2=0.52$ ,  $p=0.47$ ).

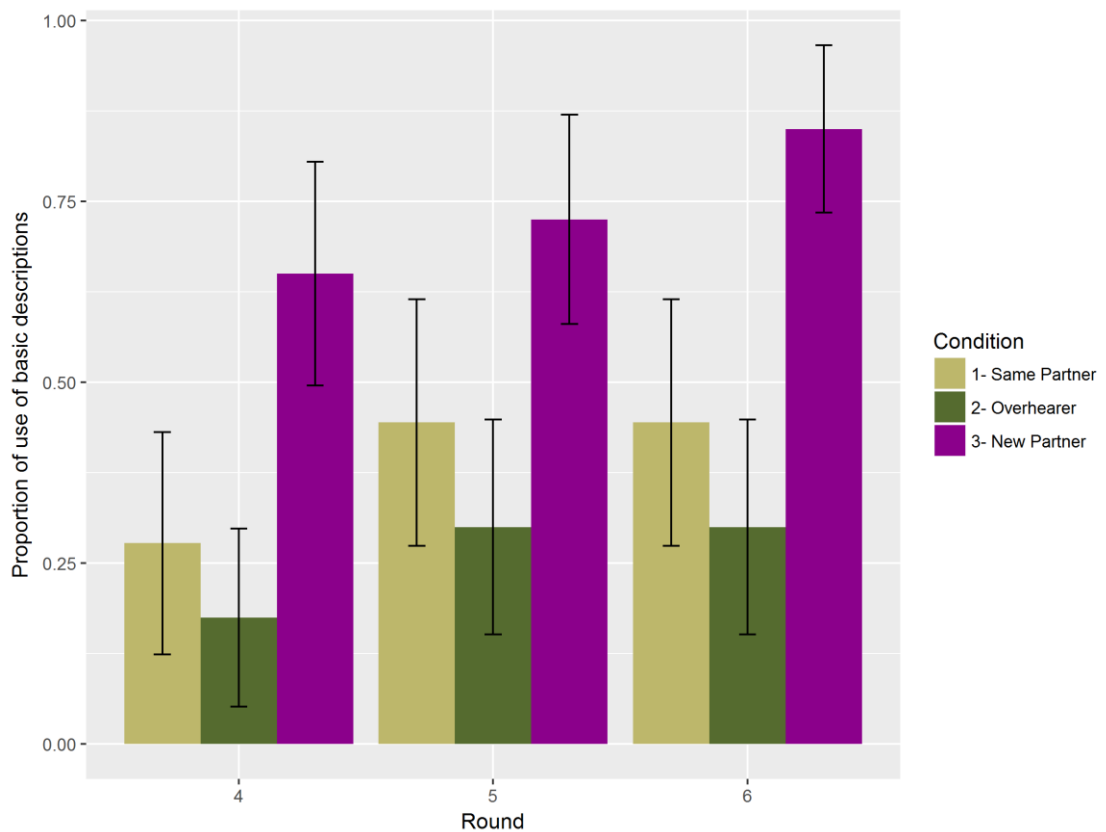


**Fig. 5.11** Proportion of use of the same description as in the previous round, Rounds 4 to 6, per Condition.

*Basic level terms.* Condition significantly predicted the use of basic terms in the first round of the second part (Round 4) (main effect of Condition  $\chi^2=9.97$ ,  $p=0.007$ ): Participants used significantly more basic level terms in the New Partner condition, compared to the Same Partner condition ( $\chi^2=4.61$ ,  $p=0.03$ ). There was no significant

difference between the Same Partner and the Overhearer conditions ( $\chi^2=0.79$ ,  $p=0.37$ ).

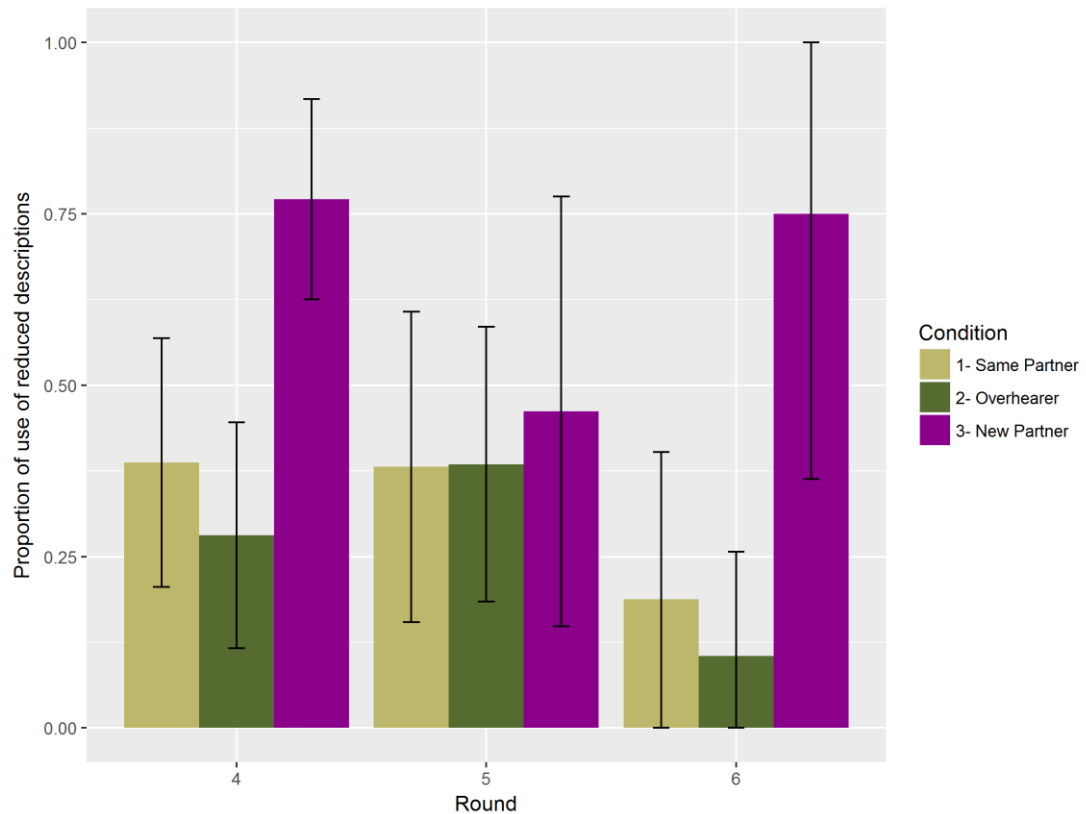
Considering Rounds 4 to 6 (see Fig. 5.12), both Condition and Round were significant predictors of the use of more basic level terms (Condition  $\chi^2=7.86$ ,  $p=0.02$ ; Round  $\chi^2=4.82$ ,  $p=0.03$ ): Participants are more likely to use basic level terms with each new round played, across Conditions; and participants in the New Partner condition are more likely to use basic level terms (compared to the Same Partner condition, marginal effect  $\chi^2=3.79$ ,  $p=0.05$ ), across Rounds. Participants in the Overhearer condition are not significantly different from participants in the Same Partner condition, across rounds ( $\chi^2=0.55$ ,  $p=0.46$ ). The interaction between Condition and Round was not significant ( $\chi^2=0.13$ ,  $p=0.94$ ).



**Fig. 5.12** Proportion of use of basic-level descriptions per rounds 4 to 6, per Condition.



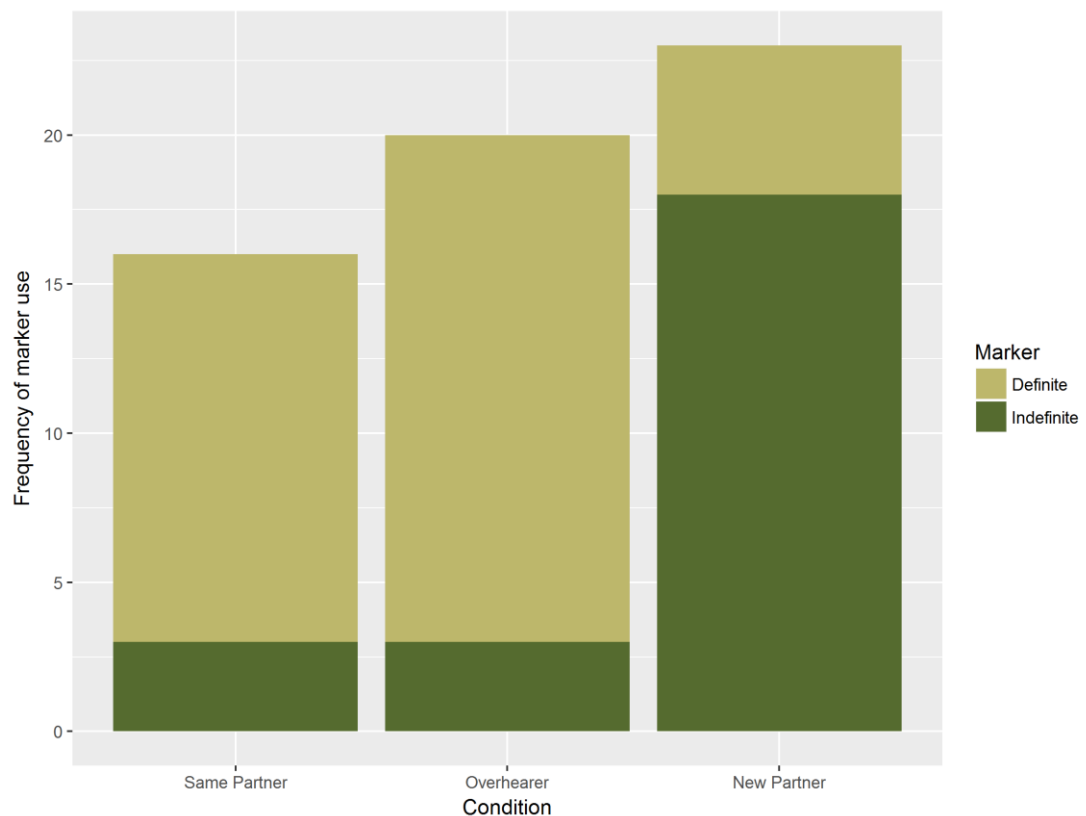
*Reduced description length.* Condition significantly predicted the use of reduced descriptions in Round 4 (main effect of Condition  $\chi^2=9.82$ ,  $p=0.007$ ). Participants reduced their descriptions more in the New Partner condition, compared to the Same Partner condition ( $\chi^2=5.17$ ,  $p=0.02$ ). There was no significant difference between the Same Partner and the Overhearer conditions ( $\chi^2=0.57$ ,  $p=0.45$ ).



**Fig. 5.13** Proportion of use of reduced descriptions in Rounds 4 to 6, per Condition. Descriptions of length 1 that were preceded by another length=1 description were not considered, as they could not possibly be reduced; therefore the proportion of Maintained descriptions only considers those in which length>1.

Considering Rounds 4 to 6 (see Fig. 5.13), Condition significantly predicted the use of more reduced descriptions ( $\chi^2=9.24$ ,  $p=0.010$ ). Participants reduced the length of their descriptions more in the New Partner condition, compared to the Same Partner condition ( $\chi^2=6.42$ ,  $p=0.01$ ), while the difference between the Overhearer and the Same Partner condition was not significant ( $\chi^2=0.03$ ,  $p=0.85$ ). Neither Round ( $\chi^2=0.02$ ,  $p=0.88$ ) nor the interaction between Condition and Round ( $\chi^2=0.16$ ,  $p=0.92$ ) were significant.

Additionally, the analysis of definite vs indefinite markers in round 4 revealed a significant difference between Conditions (proportional test,  $\chi^2=22.11$ ,  $p<0.001$ ). Considering the descriptions using markers, 81.25% of descriptions in the Same Partner condition included a definite marker (“the”), and 18.75% of descriptions included indefinite markers (“a”) (raw frequencies in Fig. 5.14). In the Overhearer condition, 85.0% of descriptions included definite markers, and 15.0% descriptions included indefinite markers. In the New Partner condition, 21.73% of descriptions included definite markers, and 78.27% of descriptions included indefinite markers.



**Fig. 5.14** Frequency of marker use in Round 4, per Condition.

*Experiments 4 and 5 comparison.* We ran an additional model comparing across experiments at Round 4. The model included both Condition and Experiment, plus their interaction; and a random intercept for Participant. We tested for the use of Basic vs Overspecific description, and the use of Reduced or Augmented descriptions vs Maintained descriptions. Differences in the use of Same description could not be measured as the full model would not converge; however, model

results for the use of Reduced descriptions should give an approximate inverse view of the Same description results.

Participants in Experiment 5 used significantly more Basic descriptions in Round 4 than participants in Experiment 4, across Conditions ( $\chi^2=17.24$ ,  $p<0.0001$ ). There was also a marginally significant interaction between Condition and Experiment, reflecting the differences in the New Partner condition between both experiments ( $\chi^2=5.00$ ,  $p=0.08$ ). Condition had no significant main effect over the use of Basic descriptions ( $\chi^2=1.84$ ,  $p=0.40$ ).

The use of Reduced descriptions was similarly affected by Experiment, with participants in Experiment 5 using significantly more reduced descriptions than participants in Experiment 4 ( $\chi^2=15.52$ ,  $p<0.0001$ ). Neither Condition ( $\chi^2=3.62$ ,  $p=0.16$ ) nor the interaction ( $\chi^2=1.90$ ,  $p=0.39$ ) were statistically significant.

### **5.3.7 Discussion**

Results of Experiment 5 show participants' choice of maintaining or abandoning entrained references in Round 4 varied according to the role their current partner had played in the past 3 rounds: In the Same Partner condition, where the partner had directly participated in previous rounds, participants kept their entrained references in higher proportion than in the New Partner condition, where the partner had not participated in previous rounds. However, in the Overhearer condition, where the partner had not participated in the references' use in previous rounds but had witnessed their use (with this fact known to the speaker), participants also kept their entrained references, in a proportion that does not differ from the Same Partner condition. We found similar results in their use of basic-level labels: participants used significantly more basic-level terms in Round 4 in the New Partner condition, compared to the Same Partner condition, and with no significant differences between the Same Partner and the Overhearer conditions. They also increased their use of basic-level labels throughout rounds 4 to 6, across conditions. Their descriptions were also significantly reduced in Round 4 in the New Partner condition, compared to the Same Partner condition, again with this factor increasing throughout rounds 4 to 6 across conditions. As in all other measures, there were no

significant differences in reference reduction between the Same Partner and the Overhearer conditions.

The difference between the New Partner and the Same Partner conditions in the use of the same terms in Round 4 as were used in Round 3 confirms that participants were switching to context-appropriate basic-level terms when they did not share a precedent with their addressee, even though these previously used terms would have been perfectly understandable in the new context (it should be equally understandable to use “*terrier*” or “*dog*” to refer to an image of a terrier in the context of other objects, even though “*dog*” should be the unmarked term). In this sense, a strong priming-exclusive explanation can be ruled out, as participants were switching to basic-level terms even when they could have maintained the precedents they were using so far. This suggests speakers combine both sources of information: their own processing effort (considering availability, production effort, etc.), and their history of interaction with their partner.

These results suggest that in a context where speakers had to produce a large number of labels, the reduction in production effort of using basic-level labels, plus the cost of switching, were lower than the formulation effort of using the same entrained precedents when these precedents were not shared with the current addressee in any way (as in the New Partner condition); considering also that the most commonly overspecified feature –colour– was not available for description. Moreover, the results in the Overhearer condition suggest that even marginal confidence in the addressee sharing a precedent is enough to tip the scale towards maintaining entrained precedents. As Exp. 4 suggested, maintaining lexical precedents can be considered the default option for speakers under low cognitive load, particularly as it can be safely assumed that the overspecified labels that were maintained were otherwise perfectly understandable by addressees both old and new.

This cognitive load assumption can be further addressed if we analyse the comparison between experiments 4 and 5 in Round 4, showing participants’ first response to the change in both context and partner. Participants in Experiment 5 use more Basic and Reduced descriptions in Round 4 across Conditions, suggesting the cognitive load of a larger item setup and the absence of the more

easily specified colour dimension had an effect on all participants independently of their partner, leading them to choose the most efficient alternative even when this meant abandoning an entrained reference.

Even though provisionality could have affected the certainty that speakers had that their conceptualisation would be accepted by an Overhearer, as they had not agreed to that conceptualisation in any way, our results indicate that this was not the case, as we found no significant differences on any measure between descriptions produced in the Same Speaker and Overhearer conditions. This suggests our speakers did not require any form of confirmation or acceptance of the conceptualisation by the Overhearer, but rather understood that the fact that they had witnessed its successful use counted as proof of their participation in its establishment.

Participants in both the Same Partner and the Overhearer conditions kept their referential precedents even when they were not only overspecific for the new context, but clearly referring to the previous context. One participant kept on calling an image of a fish "*the biggest fish*" even when there were no other fish in sight, showing that their use of the phrase could not count as a reference to the actual communicative context, but to the history of common interactions between the speakers. As scalar adjectives cannot be comprehended without a comparison class (Sedivy, 2005), it can be assumed the speaker considered the past context of the image to be the relevant comparison class.

Another speaker, after using an overspecific, previously used reference in Round 4 ("*the flopped ears dog*"), reconsidered her description and added "*the only dog*" in an additional turn immediately after. However, in the next two rounds, the speaker kept on referring to the image as "*the only dog*" (but did not do the same with "*the bee*" or "*the plant*"). This might be seen as an attempt to justify their move from an already established conventional phrase to a new one, by making their justification explicit for the addressee.

## 5.4 General discussion

Researchers have debated over many years whether previously used referential choices are repeated mainly for the processing benefit of the speaker, or, as an audience-design mechanism, for the sake of the listener, who might expect these entrained references to be maintained.

Our results showed that, in a reduced setup where the number of items to be described was low, participants retained their previously used descriptions when these became overspecific in context, even in the New Partner condition where the addressee had not been a part of the initial use of the reference (Exp. 4). However, in a larger setup where the number of items to be described was higher, and in the absence of the most easily overspecified colour dimension, participants retained their previously used descriptions only in the Same Partner and Overhearer conditions, and switched to context-appropriate basic-level descriptions when speaking to a New Partner (Exp. 5). Moreover, participants showed more adaptation to the new context in the more demanding Experiment 5 than in Experiment 4, using more Basic descriptions across the different Conditions. Both experiments showed no differences in participant behaviour between the Same Partner and the Overhearer conditions.

Taken together, our results suggest speakers combine both egocentric and allocentric sources of information, driven by a tendency to retain referential precedents if this is not detrimental to communicative success, but finally defined by the weighing of production costs of the alternatives in hand. In our setup, this weighing measured the formulation costs of maintaining overspecific labels (which should be easier to retrieve due to priming, but longer to formulate) against the switching and production costs of abandoning these labels for basic-level terms (which should have been easier to produce due to their 'unmarked' status, but costly as they involved replacing an established reference with a new label). In the small setup of Experiment 4, the potential gain in production effort of using basic-level labels instead of overspecific descriptions appears not to have outweighed the cost of abandoning an entrained conceptualisation for a new one, and therefore participants did not switch to context-appropriate alternatives (consider also that the availability of the colour dimension lowered the formulation cost of overspecific

references). In the larger setup of Experiment 5, the potential gain of using basic-level labels instead of overspecific ones appears to have been larger, as more references needed to be produced (and considering the most easily overspecified colour dimension was not available), causing more participants to switch to basic labels. However, participants switched to basic descriptions significantly more if there was no common history of use of the entrained descriptions between the speaker and the addressee. This differential behaviour suggests the common history of reference use between speakers is a factor in the weighing of production effort.

The speaker's evaluation of production effort might be, however, open to the influence of feedback. In Brennan and Clark's (1996) experiments, participants' responses were directly modified in response to verbal and non-verbal feedback, with participants being prompted to add more information if the Matcher didn't confirm her understanding in a timely fashion:

*Director:        nine is a fish [1.5 s pause]  
                      kinda green and pink  
Matcher:        done*

Participants also shortened their references if the Matcher interrupted them to signal they had already identified the intended meaning.

*Director:        number 11 is a pair –it's um, sorry, sorry, sorry, it's a fish with  
                      \*different colours\*  
Matcher:        \*yeah\* okay  
                      [Thereafter, the director used 'the fish']*

In this sense, references in Brennan and Clark's experiments were a product of both the speaker's evaluation of production effort and the addressee's communicative behaviour.

In our experiments, speakers received no direct influences over their reference choices, both as a consequence of the communication medium, and of the design, as Matchers were instructed not to comment on the Directors' choices other than to confirm or disconfirm their understanding. Moreover, as the target figures were all

simple images of common objects, neither overspecific nor basic-level labels should have compromised understanding in any way. In this sense, reference choices in our experiments can be taken to reflect more closely the speakers' evaluation of production effort, which can include their evaluation of addressees' needs.

As speakers received no feedback from Overhearers that could lead them to believe their descriptions were mutually accepted, nor interacted with them in any way (besides being co-present in the same environment, and knowing they could read the descriptions as they were produced), our results suggest the speakers' choice of referential descriptions in the second part of the experiments in the Overhearer condition was motivated exclusively by their own evaluation of the communicative situation, and not by the influence of type or quality of feedback from the addressee. This is not the same as arguing no effect of feedback at all, as the positive feedback from the original addressee in accepting the references as valid might have been enough to lead the speaker to believe in the appropriateness of his/her descriptions in that context, an evaluation that could be extended to the appropriateness of those descriptions *in general*. These results are compatible with the idea of the speaker as an opportunistic evaluator, relying on feedback when available, and on self-generated assessments of adequacy when feedback is not available (Gann & Barr, 2014).

Avoidance of other-prompted repair might also play an important role in this evaluation. Speakers try to be as informative as they need to be in order to be understood without the need to engage in an additional turn (Metzing & Brennan, 2003). In the same spirit, they seem much less likely to underspecify than to overspecify referents, as the risk of having to incur in either repair and/or an additional turn would be too high (Ferreira et al., 2005). In general, speakers seek to minimize the chance of having to be induced to a repair by the interlocutor, opting instead for 'safer' alternatives that will be less costly in terms of collaborative effort, but above all less costly in terms of their own effort. In descriptive statements, the preferred form of reference should be the one in which the speaker presents an elementary noun phrase and the addressees accept it without taking an extra turn (Clark & Wilkes-Gibbs, 1986).



However, in our experimental setup the maintenance of the descriptions that were appropriate in the first part amounts to the generation of descriptions that are overspecific for the second-part context. In this sense, an explanation in acceptability terms is not enough to account for our results, as a basic-level label would have been perfectly comprehensible in the second-part context, for both old and new addressees. Moreover, the maintenance of overspecific descriptions on both the Same Partner and the Overhearer conditions in roughly the same amounts suggests that the speakers' decision did not depend directly on feedback from the addressee that could 'bind them' to a conceptual pact. Their maintenance of overspecific descriptions is likely the outcome of an opportunistic evaluation process, where the speaker will consider the available evidence to determine 1) whether there is a risk of having to incur in repair in switching to a different alternative, and 2) the 'net saving' that could be made, in terms of production effort, by switching to basic-level labels (weighing of formulation vs switching and production costs).

While Brennan and Clark (1996) showed that speakers with new partners abandoned their previous references over the course of the 4 rounds of the second part of their experiment, we found partial evidence that speakers stopped using a linguistic precedent immediately after considering a new addressee who did not share the previous reference, switching to appropriate basic-level terms, *without any specific feedback from the addressee* that would prompt them to do so (that is, immediately in Round 4). In applying this *self-prompted adaptation* (Barr & Keysar, 2006), speakers are actively considering their partner's participation in a common history without requiring specific feedback from the partner that could serve as a communicative cue to provide less information (which was argued in Brennan & Clark (1996) to be one of the drivers of speakers' switching to basic-level terms). If the addressee is believed to have (at least) accessed the previously used reference, as in our Overhearer condition, this increases the risk that a repair or second turn might need to be employed, as priming might have pushed that reference to easier accessibility in memory than a basic-level term.

As in Gann and Barr (2014), our speakers did not switch to context-appropriate terms after feedback from their interlocutor cued them to do so, but out of their own evaluation of the communicative situation. Even though it could be suggested that

the significant effect of Round number over the use of basic-level terms reflects an effect of feedback, our confederate Matchers were instructed not to give any feedback until the description by the speaker was finished, and not to give negative feedback unless they would not understand the description if they were hearing it for the first time, which would not happen if these descriptions were overspecified. In this sense, the evaluation of the communicative context is performed solely by the speaker, with nothing more than basic confirmatory feedback from the addressee.

Taken together, our results seem to agree with a constraint-based model (Hanna & Tanenhaus, 2005), where different constraints act in parallel providing evidence that is evaluated probabilistically. These constraints can include both the evaluation of egocentric effort in the weighing of alternatives, and the consideration of an addressee's perspective, acting together in the determination of referential choices. Even though our results do not provide evidence either in favour or against conscious processing involved in this evaluation process, the immediate switch to basic-level labels in round 4 in the New Partner condition of Experiment 5 seems to suggest a decision made voluntarily by the speaker as a consequence of the change in partners, and not as part of a gradual process that might be seen as unconscious (see Branigan, 2006; Horton & Gerrig, 2005). It remains to be understood whether it is the switching process that captures conscious resources in an otherwise unconscious procedure, or if the whole process engages some degree of conscious involvement.

## **5.5 Conclusions**

The production of lexical precedents in referential tasks has been shown to be affected by the status of the addressee: if the addressee had taken part in previous uses of these references, precedents were more likely to be maintained than if the addressee was new to the communicative situation. In the two experiments reviewed in this chapter, we have shown that other factors are equally important in referential production. If the potential gain in production effort that is obtainable from switching to basic-level labels is not enough to outweigh the ease of formulation implied in maintaining primed references, speakers will not abandon their precedents, even when interacting with new partners. In the inverse situation,

speakers might abandon their precedents, but this choice will be mediated by the status of the addressee: even if the current addressee only witnessed the previous the interaction, this will be enough for speakers to retain their entrained choices in the same way as if they were talking to the partners with whom the references were first established. Our results shed light on the complexity of the weighing of production factors in reference choice, pointing towards an eminently egocentric process that is crucially defined by the history of interaction between the speaker and the addressee.

## Chapter 6. Conclusions

While the reuse of referential precedents is the most likely choice of speakers in most communicative situations, the experiments described in this thesis have shown that there are several scenarios where speakers will abandon an entrained reference, switching instead to a new alternative. Specifically, I sought to understand the influence of two factors over this referential adaptation process: the perceptual context, and the participant role of speakers and hearers. The perceptual context refers to all the elements that surround the interaction to which the speaker (and usually the hearer) has direct (i.e. non-linguistic) access, including the task context. Depending on the specific communicative situation, the perceptual context might constitute a sort of 'pre-linguistic common ground' (Wilkes-Gibbs & Clark, 1992) that the interlocutors can consider as shared when planning and producing an utterance.

The participant role of speakers and hearers refers to the degree of centrality and/or accountability of the individuals involved in the interaction. While the speaker and the addressee are central to the interaction and can be held accountable for the understanding of all references that have been mutually accepted in the conversation, side-participants who are not directly addressed might not share the same degree of understanding, while overhearers who did not directly participate in the interaction nor were acknowledged as participants have been described as not sharing the same discourse record as acknowledged participants (Branigan, 2006).

The experiments presented in this thesis have shown that both factors influence referential adaptation. The influence of context was not only significant in the speakers' initial evaluation of the communicative situation, but also as a consequence of their production and comprehension experience, with participants changing their

reference choices after having the experience of using an alternative description in context. This experience effect suggests speakers are constantly evaluating the fit between their linguistic production and the context where this production will be understood. Participant role, on the other hand, was shown to influence speakers' referential choices only in its most contrasting version –that of a known interlocutor with whom the references were previously established, vs a new interlocutor with no common interaction history with the speaker.

However, while our results show both factors influence referential adaptation, they also suggest this influence is subtle and immersed in a complex dynamical process. Even though the perceptual context might be assumed to exert a strong pressure over speakers' choices, the experiments show that this influence might be overridden by production factors or communicative demands. In this sense, these results call for careful consideration of the experimental setup and stimuli used by experiments looking to address this issue, as what appears as a stable response in one context might not be so certain in a slightly different setting (see, for example, the differences between Experiments 4 and 5). Similarly, the results of our overhearer conditions (Experiments 3, 4, and 5), in which no significant differences were found between participants talking to or performing as overhearers compared to main speakers or addressees, suggest differences between participant roles might appear only in specific conditions, particularly in situations where the difficulty of the task gives extra relevance to any minor advantage in individual terms, and might be unnoticeable in simpler referential tasks.

As a whole, the research summarised in this thesis was successful in describing referential adaptation –a neglected aspect in the literature on reference production– and establishing two of the main factors that influence its appearance. Moreover, it contributed to the current discussion on the articulation of speaker-centred and audience-centred factors in reference production, showing how both influences are weighed in an experimental task, and how minor differences in contextual factors can bring forth significant differences in speakers' choices.

## 6.1 Summary

This thesis proposed an exploration of the factors of the context and the interaction that would potentially influence speakers' reference choice, particularly the perceptual context and the participant role of the individuals involved in the interaction. I designed five experiments that would address these factors, comparing speakers in different interactive and non-interactive settings in their reuse or adaptation of reference choices.

Chapter 2 presented a review of the literature surrounding the reference production process, going from the factors that influence the initial selection of a conceptual perspective for a reference, to the consequences of the reuse of a reference both locally in an interacting pair, and across communities of speakers. The chapter discussed how repetition can be considered the default option in most reference use, as it facilitates both production and comprehension, easing the cognitive load of the speaker due to routinization of its local meaning. Moreover, it showed how the reuse of references can be linked to the emergence of linguistic conventions, and on a broader scale, to language change, as each instance of reuse adds to the strength of the emergent convention.

Experiments 1 to 3 in Chapters 3 and 4 used the maze game (Garrod & Anderson, 1987) to address the influence of the perceptual context and the interactivity of the setting in participants' reference choices. The experiments share a pattern of adaptation to the repetition of the task, with speakers using more abstract descriptions—which are more efficient when facing the need to repeatedly refer to positions in one or more mazes—as they played more rounds of the game. Experiment 1 showed that speakers' initial description choice was influenced by the layout of the maze, with more concrete and idiosyncratic descriptions being preferentially used in irregular mazes, and more abstract descriptions mainly used in regular mazes. Moreover, the likelihood of participants maintaining this initial choice, or adapting instead to the repetition of the task, switching to abstract descriptions, was linked to the possibility of interaction in their experimental condition: Participants working in interactive pairs adapted to the circumstances of their interaction, using more abstract descriptions as they played; while participants working individually maintained their initial reference choices.

In Experiments 2 and 3, I explored specific aspects of the task context that could influence referential adaptation. Experiment 2 addressed the interaction between context change and adaptation, presenting pairs and individual participants with either the same irregular maze across rounds, or different irregular or regular mazes. In the task, all participants started playing in the same maze layout, facing different context conditions after completing the first round. The results revealed this context manipulation did not influence participants' referential adaptation; however, participants across conditions adapted to the repetition of the task by using progressively more abstract descriptions as they played the game. Experiment 3 addressed the relationship between participant role in the interaction and referential adaptation, by distributing participants into different participant role conditions where they could either play with the same partner throughout the task, or with a different partner in the second half of the experiment. While the different participant roles did not significantly influence participants' referential adaptation, the results showed that task experience was again a factor in the use of more abstract descriptions.

Experiments 4 and 5, in Chapter 5, looked at the production of referential descriptions in a picture sorting task, where target images would appear along with other images of the same category in the first part of the task (e.g. a target dog with two other dogs in the set), and alone in their category in the second part of the task (e.g. the same dog with no other dogs in the set). As the references used in the first part became overspecific in the second part, participants could either maintain the previously used references, or switch to basic-level labels. Moreover, participants were distributed into three different conditions that determined their partner's identity in the second part of the game: a Same Partner condition, where they maintained the same partner as before; an Overhearer condition, where they went on to play with the person who had been witnessing their interaction in the first half of the game; and a New Partner condition. Experiment 4 used a small set of colour pictures, and showed that participants maintained their entrained references even if they were interacting with a new partner who did not share the past history of use of these references. Experiment 5 used a larger image set of monochrome pictures, showing that participants again chose to maintain their referential precedents *unless* they were interacting with a new partner who did not share these references, in which case they switched to basic-level labels. Across the two experiments, participants were more likely to maintain referential precedents even when the context change made them overspecific;

however, the results of Experiment 5 show speakers are willing to abandon referential precedents in the right conditions, particularly if the formulation cost of these precedents is high.

## 6.2 Discussion

The production of a reference requires that the speaker performs a complex (but not necessarily conscious) evaluation of the contextual and historical circumstances of the interaction, in order to find a balance between production effort, informativeness, and the needs of the addressee. The main aim of this thesis was establishing under which conditions speakers would be likely to abandon a referential precedent –which can be considered the ‘default’ choice, as maintaining a precedent is regularly easier for both speaker and addressee to process than switching to a different option. I have conceptualised this process as *referential adaptation*, as the speakers abandoning a previously used reference are adapting to specific aspects of their circumstances of use by switching to what is perceived as a ‘better’ alternative in those circumstances. This preference for ‘better’ alternatives, which can be more efficient, or easier to produce, or easier to understand for an addressee, can be linked to the dynamics of language change in a population, where signs are selected according to their past use and functionality in context (Tamariz et al., 2014).

Taken together, the experiments described in this thesis point to a tendency to retain previously used references that can be overcome in the right circumstances, particularly by speakers in interaction with a partner or in a demanding context. Even though Experiments 2 and 3 aimed at establishing which aspects of the interaction were particularly relevant for this adaptation process, the experimental manipulations did not succeed at identifying any differences in participants’ performance in individual vs. interactive settings, nor in participants with a central vs. a more peripheral role in the interaction. In these experiments, the changes in the perceptual context pushed all participants to adapt by switching to abstract reference schemes, which were a better fit for a variable scenario.

On the other hand, Experiments 4 and 5 showed that, in a less demanding context, participants only adapted their references when the formulation costs of the



references they were previously using were high, that is, in the presence of a large number of items and without the most easily overspecified colour dimension; and when there was no common history of use of these references with the current partner. When the formulation cost of the precedents was lower, participants retained them, suggesting a conservative tendency defines the default reference choice.

In general terms, these results seem to point towards a prominent role for egocentric processes that are crucially open to consideration of the addressee's identity and needs. In this sense, it could be suggested that referential decisions are weighed in a formula that assigns more value to speaker-internal processes. However, this research did not present speakers with any situation where their decisions could be radically detrimental to communication, such as underspecifying a referent; and in that sense it should be noted that a *least collaborative effort* framework might still be used to explain these results. As both maintaining referential precedents and switching to a new alternative were valid options in terms of communicative effectiveness (i.e. both were able to convey the intended message, even if not with the same efficiency), speakers could prioritise lowering their production costs at a likely low cost for the addressee.

### **6.3 Limitations and implications for future research**

This thesis used two different experimental paradigms in an attempt to establish the conditions in which referential adaptation can operate. While all of the experiments described in this work showed some degree of referential adaptation, not all the manipulations that we set out to test specific aspects of this problem were successful in finding significant differences in speakers' choices. On one hand, the variability of initial choices between individuals in the same condition in the maze game studies, which cannot be attributed to any experimental manipulation, had an impact on the experiments' results, as these initial choices act as precedents for all subsequent descriptions. While Experiment 1 showed a consistent pattern of different description scheme choices in different maze layouts, in Experiments 2 and 3 the same initial maze layout was described using different description schemes. Even though this issue is informative on its own, as it suggests the impact of the perceptual context is subject to individual variation, the experimental setup I designed failed to consider

that the influence of the experimental conditions could be masked by these initial differences.

In general terms, the experimental methods used allowed us to look only at the linguistic level of reference choice, that is, at the actual linguistic production of the speakers, but not at other more subtle cues of the processing and weighing of factors that could have influenced their choice. In this sense, it is likely that additional experimental methods –like eye-tracking, or timing of articulation in spoken dialogue– would be needed to gain a better understanding of the processes that lead to the actual articulation of a referential choice.

Regarding the influence of the experiment design and methods, the results suggest significant differences in results might emerge from apparently minor differences in implementation, such as those between the setups of Experiments 4 and 5. This outcome points to the need to diversify the range of tests that are used in language production research, as subtle factors that are not considered or controlled might be influencing the patterns of results. In this sense, it is particularly interesting to notice how both the availability of perceptual dimensions and the number of descriptions to be produced in referential tasks could potentially radically alter speakers' evaluation of their referential options.

One of the areas where it could be suggested that the use of different methods might allow for the exploration of more subtle differences is participant role. This thesis showed that participant role differences were only associated to different referential outcomes in their most extreme form, that of a previously known partner with whom the references were initially established, versus a new partner who had no knowledge of these references nor had participated in the interaction before. More subtle differences, like that of a known partner versus an overhearer who was physically present in the scene but not involved in the interaction, were not associated to differences in reference choice. However, previous research had suggested the degree of involvement in the interaction and the access of individuals to task information should affect the linguistic choices of participants. Considering this, I believe looking at this issue with additional experimental methods, or with a more challenging task, might help defining the extent to which these different roles affect reference choice. Particularly, the use of more complex stimuli where a common

interpretation cannot be easily assumed might allow for a better understanding of subtle differences in the speaker's evaluation of their partner's communicative needs.

## References

- Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, 21(12), 1903–9.
- Anderson, A., & Garrod, S. C. (1987). The dynamics of referential meaning in spontaneous conversation: Some preliminary studies. In R. Reilly (Ed.) *Communication failure in dialogue and discourse*, 161–183. Amsterdam: Elsevier.
- Anderson, A., Garrod, S. C., & Sanford, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *The Quarterly Journal of Experimental Psychology Section A*, 35(3), 427–440.
- Ariel, M. (1990). *Accessing NP antecedents*. London: Croom Helm.
- Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23(4), 495–527.
- Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics*, 43(1), 361–374.
- Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, 27(2), 195-225.
- Bard, E. G. & Aylett, M. P. (2004). Referential form, word duration, and modeling the listener in spoken dialogue. In J. Trueswell & M. Tanenhaus (Eds.) *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*, 173-191. Cambridge, MA: The MIT Press.
- Baronchelli, A., Felici, M., Loreto, V., Caglioti, E., & Steels, L. (2006). Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(06), P06014.
- Barr, D. J. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science*, 28(6), 937–962.
- Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, 46, 391-418.
- Barr, D. J., & Keysar, B. (2006). Perspective taking and the coordination of meaning in language use. In M. Traxler & M. Gernsbacher (Eds.) *Handbook of Psycholinguistics (Second Edition)*, 901-938. London: Academic Press.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and*

*Language*, 68(3).

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Bates, E., Masling, M., & Kintsch, W. (1978). Recognition memory for aspects of dialogue. *Journal of Experimental Psychology: Human Learning and Memory*, 4(3), 187.
- Behnel, M., Cummins, C., Sichelschmidt, L., & De Ruiter, J. (2013). Priming and conceptual pacts in overhearers' adoption of referring expressions. *Proceedings of the Cognitive Science Society*, 35(35). Retrieved from <http://escholarship.org/uc/item/2k5493g0#page-1>
- Bock, J. K., & Warren, R. K. (1985). *Conceptual accessibility and syntactic structure in sentence formulation*. *Cognition* (Vol. 21). Retrieved from <http://www.linguisticsnetwork.com/wp-content/uploads/Bock-Warren-1986.pdf>
- Branigan, H. (2006). Perspectives on multi-party dialogue. *Research on Language and Computation*, 4(2-3), 153-177.
- Branigan, H. P., Catchpole, C. M., & Pickering, M. J. (2011). What makes dialogues easy to understand? *Language and Cognitive Processes*, 26(10), 1667–1686.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(B), 13–25.
- Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, 104(2), 163–197.
- Branigan, H. P., Pickering, M. J., Pearson, J., Mclean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, 121, 41–57.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–1493.
- Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1(2), 274–291.
- Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue and Discourse*, 2, 11–33.
- Brown-Schmidt, S., & Konopka, A. E. (2011). Experimental approaches to referential domains and the on-line processing of referring expressions in unscripted conversation. *Information*, 2(2), 302–326.

- Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, 27(1), 62-89.
- Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as contexts in conversation. *Psychology of Learning and Motivation*, 62, 59–99.
- Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, 15(1), 73.
- Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *PNAS*, 112(7), 1989–1994.
- Chartrand, T., & Bargh, J. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893.
- Chartrand, T. L., & van Baaren, R. (2009). Human mimicry. *Advances in Experimental Social Psychology*, 41, 219–274.
- Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, 767-811.
- Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. *Mechanisms of language acquisition*, 1, 33.
- Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. *Cognition*, 64, 1–37.
- Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. *Attention and performance IX*, 313-330. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Clark, H. H. (1985). Language use and language users. In Lindzey, G., & Aronson, E. (Eds.). *Handbook of social psychology* (3rd ed.). New York: Hillsdale, N.J.: Random House.
- Clark, H. H. (1996). *Using language*. Cambridge, MA: Cambridge University Press.
- Clark, H., & Brennan, S. (1991). Grounding in communication. In L. Resnik, J. Levine, and S. Teasley (Eds.) *Perspectives on Socially Shared Cognition*, 222–233. Washington, DC: American Psychological Association.
- Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1), 62-81.
- Clark, H. & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B.L. Webber, & I. A. Sag. (Eds.), *Elements of discourse understanding*, 10-63. Cambridge, MA: Cambridge University Press.
- Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, 2(1), 19-41.
- Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*,

13(2), 259–294.

- Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, 22, 245–258.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.
- Clarke, A. D. F., Elsner, M., & Rohde, H. (2013). Where's Wally: the influence of visual salience on referring expression generation. *Frontiers in Psychology*, 4, 329.
- Coco, M. I., Dale, R., & Keller, F. (2018). Performance in a collaborative search task: The role of feedback and alignment. *Topics in Cognitive Science*, 10(1), 55–79.
- Croft, W. (2013). Evolution: Language use and the evolution of languages. In: Binder PM., Smith K. (Eds.) *The Language Phenomenon*. The Frontiers Collection. Berlin, Heidelberg: Springer.
- Cruse, D. A. (1977). The pragmatics of lexical specificity. *Journal of Linguistics*, 13(2), 153-164.
- Dale, R., Galati, A., Alviar, C., Contreras Kallens, P., Ramirez-Aristizabal, A., Tabatabaeian, M., & Vinson, D. W. (2018). Interacting timescales in perspective-taking. *Frontiers in Psychology*, 9, 1278.
- Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, 39(4), 667–710.
- Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science*, 40(1), 172–201.
- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19(10), 603–615.
- Engelhardt, P., Bailey, K., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, 54(4), 554–573.
- Engelhardt, P., Demiral, B., & Ferreira, F. (2011). Over-specified referring expressions impair comprehension: An ERP study. *Brain and Cognition*, 77(2), 304–314.
- Fay, N., & Ellison, T. M. (2013). The cultural evolution of human communication

- systems in different sized populations: Usability trumps learnability. *PLoS ONE*, 8(8).
- Fay, N., Ellison, T. M., & Garrod, S. (2014). Iconicity: From sign to system in human communication and language. *Pragmatics & Cognition*, 22(2), 244–263.
- Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B-Biological Sciences: Biological Sciences*, 363(1509), 3553–3561.
- Fay, N., Walker, B., Swoboda, N., & Garrod, S. (2018). How to create shared symbols. *Cognitive Science*, 42, 241–269.
- Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*, 91, 158-180.
- Ferreira, V. S., & Dell, G. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40(4), 296–340.
- Ferreira, V. S., Slevc, R., & Rogers, E. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, 96(3), 263–284.
- Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes*, 27(1), 35–53.
- Fox Tree, J. E., & Mayer, S. A. (2008). Overhearing single and multiple perspectives. *Discourse Processes*, 45(2), 160–179.
- Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: quantifying the benefits of linguistic coordination. *Psychological Science*, 23(8), 931-939.
- Fusaroli, R., Raczaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32(1), 147–157.
- Fusaroli, R., & Tylén, K. (2012). Carving language for social coordination: A dynamical approach. *Interaction Studies*, 13(1), 103–124.
- Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive Science*, 40(1), 145–171.
- Fusaroli, R., Tylén, K., Garly, K., Steensig, J., Christiansen, M. H., & Dingemanse, M. (2017). Measures and mechanisms of common ground: backchannels, conversational repair, and interactive alignment in free and task-oriented social interactions. In the *39th Annual Conference of the Cognitive Science Society (CogSci 2017)* (pp. 2055-2060). Cognitive Science Society.



- Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, 62(3), 378.
- Fukumura, K., & van Gompel, R. P. (2012). Producing pronouns and definite noun phrases: Do speakers use the addressee's discourse model?. *Cognitive Science*, 36(7), 1289-1311.
- Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, 29(5), 737–767.
- Galantucci, B., Rhodes, T., & Kroos, C. (2006). Rapidity of fading and the emergence of duality of patterning. In *The Evolution of Language: Proceedings of the 6th International Conference (EVOLANG6)*, Rome, Italy, 12-15 April 2006.
- Galantucci, B., & Sebanz, N. (2009). Joint action: Current perspectives. *Topics in Cognitive Science*, 1(2), 255–259.
- Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee?. *Journal of Memory and Language*, 62(1), 35-51.
- Gann, T. M., & Barr, D. J. (2014). Speaking from experience: Audience design as expert performance. *Language, Cognition and Neuroscience*, 29(6), 744-760.
- Garrod, S. (2011) Referential processing in monologue and dialogue with and without access to real-world referents. In: Gibson, E., & Pearlmutter, N. (Eds.) *The processing and acquisition of reference*. Cambridge, Mass. ; London: MIT Press.
- Garrod, S., Anderson, A., & Sanford, A. J. (1984). Semantic negotiation and the dynamics of conversational meaning. *Glasgow Psychology Technical Report*, (1).
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2), 181–218.
- Garrod, S., & Clark, A. (1993). The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes*, 8, 37–41.
- Garrod, S., & Doherty, G. (1994). Conversation, coordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53, 181–215.
- Garrod, S., Fay, N., Lee, J., Oberlander, J., & Macleod, T. (2007). Foundations of representation: where might graphical symbol systems come from? *Cognitive*

- Science*, 31(6), 961–987.
- Garrod, S., & Pickering, M. J. (2007). *Alignment in dialogue*. Oxford: Oxford University Press.
- Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292–304.
- Garrod S., & Pickering M.J. (2013) Dialogue: Interactive alignment and its implications for language learning and language change. In: Binder PM., Smith K. (eds) *The Language Phenomenon*. The Frontiers Collection. Berlin, Heidelberg: Springer.
- Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of Accommodation* (pp. 1–68). Cambridge: Cambridge University Press.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press.
- 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(3), 843.
- Goudbeek, M., & Kraemer, E. (2012). Alignment in interactive reference production: Content planning, modifier ordering, and referential overspecification. *Topics in Cognitive Science*, 4(2), 269-289.
- Grice, H.P. (1975). Logic and conversation. In Cole et al. *Syntax and semantics 3: Speech arts* (pp. 41-58). Retrieved from <http://www.ucl.ac.uk/ls/studypacks/Grice-Logic.pdf>
- Grodner, D., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In N. Pearlmuter & E. Gibson (Eds.) *The processing and acquisition of reference*. Cambridge, MA: Cambridge University Press.
- Hanna, J., Tanenhaus, M., & Trueswell, J. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49(1), 43–61.
- Hanna, J. & Tanenhaus, M. (2005). The use of perspective during referential interpretation. In J. Trueswell & M. Tanenhaus (Eds.) *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*, 133-152. Cambridge, MA: The MIT Press.

- Haselow, A. (2018). Language change from a psycholinguistic perspective: The long-term effects of frequency on language processing. *Language Sciences*, 68, 56–77.
- Hawkins, R. X., Frank, M. C., & Goodman, N. D. (2017). Convention-formation in iterated reference games. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. Cognitive Science Society.
- Healey, P. G. T., Mills, G. J., Eshghi, A., & Howes, C. (2018). Running repairs: Coordinating meaning in dialogue. *Topics in Cognitive Science*, 10(2), 367–388.
- Healey, P. G. T., Mills, G., & Eshghi, A. (2016). Co-ordinating understanding: The importance of negative evidence. *International Journal of Psychology* 51, 1040.
- Healey, P. G. T., & Mills, G. (2006). Participation, precedence and co-ordination in dialogue. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (Vol. 320). Vancouver: Cognitive Science Society.
- Healey, P. G. T., & Mills, G. J. (2009, January). A Dialogue Experimentation Toolkit. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 31, No. 31).
- Healey, P. G. T., & Mills, G. J. (submitted) A Dialogue Experimentation Toolkit. Retrieved from <http://homepages.inf.ed.ac.uk/gmills/MillsHealey2013Submission.pdf>
- Healey, P. G. T. (1997). Expertise or expert-ese?: The emergence of task-oriented sub-languages. *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, 301–306.
- Healey, P. G. T. (2008). Interactive misalignment: The role of repair in the development of group sub-languages. *Language in Flux*, 13–39.
- Healey, P. G. T., Purver, M., & Howes, C. (2014). Divergence in dialogue. *PLoS One*, 9(2), e98598.
- Herring, S. C. (1999, January). Interactional coherence in CMC. HICSS-32. *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences* (pp. 13-pp). IEEE.
- Heyes, C. (2011). Automatic imitation. *Psychological Bulletin*, 137(3), 463.
- Horton, W. S. (2008). A memory-based approach to common ground and audience design. In I. Kecskes & J. Mey (Eds.) *Intention, common ground, and the egocentric speaker-hearer*, 189-222. Berlin: Mouton de Gruyter.
- Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design:

- knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, 47, 589–606.
- Horton, W. S., & Gerrig, R. J. (2005). Conversational common ground and memory processes in language production. *Discourse Processes*, 40(1), 1–35.
- Horton, W. S., & Gerrig, R. J. (2016). Revisiting the memory-based processing approach to common ground. *Topics in Cognitive Science*, 8(4), 780–795.
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59(1), 91–117.
- 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(1), 113–126.
- Hupet, M., & Chantraine, Y. (1992). Changes in repeated references: Collaboration or repetition effects?. *Journal of Psycholinguistic Research*, 21(6), 485–496.
- 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(4), 449–462.
- Ibarra, A., & Tanenhaus, M. K. (2016). The flexibility of conceptual pacts: Referring expressions dynamically shift to accommodate new conceptualizations. *Frontiers in Psychology*, 7, 561.
- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130298–20130298.
- Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116(1), 26.
- Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, 49(10), 1295–1306.
- Keysar, B. (1997). Unconfounding common ground. *Discourse Processes*, 24(2-3), 253–270.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32–38.
- Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, 39.
- Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language

- use: Insights from a processing approach. *Current Directions in Psychological Science*, 7(2), 46-49.
- Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness monitoring and ambiguity. *Psychological Science*, 13(3), 207–212.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89(1), 25–41.
- Keysar, B., & Barr, D. J. (2005). Coordination of action and belief in communication. In J. Trueswell & M. Tanenhaus (Eds.) *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*, 71-94. Cambridge, MA: The MIT Press.
- Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin & Review*, 1590–1599.
- Knutsen, D., Le Bigot, L., & Ros, C. (2017). Explicit feedback from users attenuates memory biases in human-system dialogue. *International Journal of Human-Computer Studies*, 97, 77–87.
- Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1(1-12), 113-114.
- Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4(3), 343.
- Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language*, 83, 1–19.
- Kronmüller, E., Noveck, I., Rivera, N., Jaume-Guazzini, F., & Barr, D. (2017). The positive side of a negative reference: the delay between linguistic processing and common ground. *Royal Society Open Science*, 4.
- Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, 20(1), 54-72.
- Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, 14(4), 334–339.
- Lewis, D. K. (2002). *Convention : a philosophical study*. Oxford, UK: Blackwell Publishing. (Originally published 1969).
- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind

- to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, 46(3), 551–556.
- Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication Is synchronized. *Cognitive Science*, 36(8), 1404–1426.
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS ONE*, 5(1), e8559.
- Lupyan, G., & Dale, R. A. (2015). The role of adaptation in understanding linguistic diversity. *Language structure and environment: Social, cultural, and natural factors*, 287-16.
- Lupyan, G., & Dale, R. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends in Cognitive Sciences*, 20(9), 649–660.
- McClung, J. S., & Reicher, S. D. (2018). Representing other minds: Mental state reference is moderated by group membership. *Journal of Experimental Social Psychology*, 76, 385–392.
- 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(2), 201–213.
- Mills, G. J., & Healey, P. G. (2006). Clarifying spatial descriptions: Local and global effects on semantic co-ordination. *Proc. Workshop on Semantics and Pragmatics of Dialogue*, 122–129.
- Mills, G. J. (2014). Dialogue in joint activity: Complementarity, convergence and conventionalization. *New Ideas in Psychology*, 32(1), 158–173.
- Misyak, J., Noguchi, T., & Chater, N. (2016). Instantaneous conventions: The emergence of flexible communicative signals. *Psychological Science*, 0956797616661199.
- Oben, B., & Brone, G. (2016). Explaining interactive alignment: A multimodal and multifactorial account. *Journal of Pragmatics*, 104, 32–51.
- Paraboni, I., & van Deemter, K. (2014). Reference and the facilitation of search in spatial domains. *Language, Cognition and Neuroscience*, 29(8), 1002–1017.
- Paraboni, I., van Deemter, K., & Masthoff, J. (2007). Generating referring expressions: Making referents easy to identify. *Computational Linguistics*, 33(2), 229–254.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382–2393.

- Paukner, A., Suomi, S. J., Visalberghi, E., & Ferrari, P. F. (2009). Capuchin monkeys display affiliation toward humans who imitate them. *Science*, 325(5942), 880–3.
- Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, 27(1), 89-110.
- Perfors, A., & Navarro, D. J. (2014). Language evolution can be shaped by the structure of the world. *Cognitive Science*, 38(4), 775–793.
- Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, 1(DEC), 1–15.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, 27(2), 169-190; discussion 190-226.
- Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4(2–3), 203–228.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105–110.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, 36(4), 329–47.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372.
- Rogers, S. L., Fay, N., & Maybery, M. (2013). Audience design through social interaction during group discussion. *PLoS ONE*, 8(2), e57211.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382-439.
- Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, 7, 153.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2), 361-382.
- Schober, F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21, 211–232.
- Sedivy, J. (1999). Examining the discourse-based properties of adjectives in on-line

- semantic processing. In *Twelfth Annual CUNY Conference on Human Sentence Processing*, New York.
- Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, 32(1), 3–23.
- Sedivy, J. (2005). Evaluating explanations for referential context effects: Evidence for gricean mechanisms in online language interpretation. In J. Trueswell & M. Tanenhaus (Eds.) *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*, 173-191. Cambridge, MA: The MIT Press.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71(2), 109-147.
- Senay, I., & Keysar, B. (2009). Keeping track of speaker's perspective: The role of social identity. *Discourse Processes*, 46(5), 401–425.
- Shintel, H., & Keysar, B. (2007). You said it before and you'll say it again: Expectations of consistency in communication. *Journal of Experimental Psychology: Learning Memory and Cognition*.
- Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science*, 1(2), 260–273.
- Shirouzu, H., Miyake, N., & Masukawa, H. (2002). Cognitively active externalization for situated reflection. *Cognitive Science*, 26(4), 469–501.
- Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 326–332.
- Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). afex: Analysis of Factorial Experiments. R package version 0.16-1. <https://CRAN.R-project.org/package=afex>
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 592.
- Smith, K., Fehér, O., & Ritt, N. (2014). Eliminating unpredictable linguistic variation through interaction. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, 1461–1466.
- Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, 116(3), 444–449.



- Stalnaker, R. (1978). Assertion. *Syntax and Semantics* 9(3), 15-22 .
- Tamariz, M., Ellison, T. M., Barr, D. J., & Fay, N. (2014). Cultural selection drives the evolution of human communication systems. *Philosophical Transactions of the Royal Society B-Biological Sciences: Biological Sciences*, 281(June).
- Tamariz, M., Roberts, S. G., Martínez, J. I., & Santiago, J. (2018). The Interactive origin of iconicity. *Cognitive Science*, 42(1), 334–349.
- Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder?. *Cognitive Psychology*, 23(3), 457-482.
- Tannen, D. (2007). *Talking voices: Repetition, dialogue, and imagery in conversational discourse (Studies in Interactional Sociolinguistics)*. Cambridge: Cambridge University Press.
- Tarenskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: salience, absoluteness, and consistency. *Frontiers in Psychology*, 6, 1703.
- Tolins, J., & Tree, J. E. F. (2014). Addressee backchannels steer narrative development. *Journal of Pragmatics*, 70, 152-164.
- Tollefsen, D., & Dale, R. (2012). Naturalizing joint action: A process-based approach. *Philosophical Psychology*, 25(3), 385–407.
- Trueswell, J. C., & Tanenhaus, M. K. (2005). *Approaches to studying world-situated language use : bridging the language-as-product and language-as-action traditions*. Cambridge, MA: The MIT Press.
- van Baaren, R., Holland, R., Steenaert, B., & van Knippenberg, A. (2003). Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology*, 39(4), 393–398.
- van Deemter, K., Gatt, A., van Gompel, R. P. G., & Krahmer, E. (2012). Toward a computational psycholinguistics of reference production. *Topics in Cognitive Science*, 4(2), 166–183.
- Van Der Wege, M. M. (2009). Lexical entrainment and lexical differentiation in reference phrase choice. *Journal of Memory and Language*, 60(4), 448–463.
- Viethen, J., Goudbeek, M., & Krahmer, E. (2012, January). The impact of colour difference and colour codability on reference production. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34, No. 34).
- Vogels, J., Krahmer, E., & Maes, A. (2013). Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production. *Language and Cognitive Processes*, 28.

- Wardlow Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! *Psychological Science*, 17(4), 273–277.
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31(2), 183-194.
- Wilkes-Gibbs, D. (1995). Coherence in collaboration: Some examples from conversation. *Typological Studies in Language*, 31, 239-267.
- Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131(3), 460.
- Winters, J., Kirby, S., & Smith, K. (2018). Contextual predictability shapes signal autonomy. *Cognition*, 176(March), 15–30.
- Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(4), 919.
- Yoon, S. O., Koh, S., & Brown-Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, 19(4), 699–707.