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.
ADJUSTING LINGUISTICALLY TO OTHERS: THE ROLE OF SOCIAL CONTEXT IN LEXICAL CHOICES AND SPATIAL LANGUAGE

Alessia Tosi

PhD Psychology

The University of Edinburgh

2016
DECLARATION

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Alessia Tosi

Edinburgh, August 25th, 2016
The human brain is highly sensitive to social information and so is our language production system: people adjust not just what they say but also how they say it in response to the social context. For instance, we are sensitive to the presence of others, and our interactional expectations and goals affect how we individually choose to talk about and refer to things.

This thesis is an investigation of the social factors that might lead speakers to adapt linguistically to others. The question of linguistic adaptation is conceived and addressed at two levels: as lexical convergence (i.e., interlocutors coordinating their lexical choices with each other), and as spatial perspective taking in language use (i.e., speakers abandoning their self perspective in favour of another's when verbally locating objects in space).

What motivated my research was two-fold. First, I aimed to contribute to the understanding of the interplay between the automatic cognitive accounts and the strategic social accounts of linguistic convergence. At the same time, I wanted to explore new analytical tools for the investigation of interpersonal coordination in conversation (cross-recurrence quantification analysis (CRQA)). Second, there are conflicting explanations as to why people often abandon their self spatial perspective when another person is present in the environment. I aimed to clarify this by bringing together insights from different research fields: spatial language production, spatial cognition, joint attention and joint action.

A first set of experiments investigated the effects of speakers' deceptive goals on lexical convergence. Given the extensive evidence that one interlocutor's choices of words shapes another's during collaborative interaction, would we still observe this coordination of linguistic behaviour under conditions of no coordination of intents? In two novel interactive priming paradigms, half of the participants deceived their naïve partner in a detective game (Experiment 1) or a picture naming/matching task (Experiment 2-3) in order to jeopardise their partner's performance in resolving the crime or in a related memory task. Crucially, participants were primed by their partner with suitable-yet-unusual names for objects. I did not find any consistent evidence that deceiving led to a different degree of lexical convergence between deceivers and deceived than between truthful interlocutors.
I then explored possibilities and challenges of the use of cross-recurrence quantification analysis (CRQA) (a new analytical tool borrowed from dynamical systems) for the study of lexical convergence in conversation. I applied CRQA in Experiment 4, where I focused on the strategic social accounts of linguistic convergence and investigated whether speakers' tendency to match their interlocutors' lexical choices depended on the social impression that they formed of each other in a previous interaction, and whether this tendency was further modulated by the interactional goal. I developed a novel two-stage paradigm: pairs of participants first experienced a collectivist or an individualistic co-player in an economic decision game (in reality, a pre-set computer programme) and then engaged in a discussion of a survival scenario (this time with the real other) divided in an open-ended vs. joint-goal driven part. I found no evidence that the social impression of their interlocutor affected speakers' degree of lexical convergence. Greater convergence was observed in the joint-goal dialogues, replicating previous findings at syntactic level.

Experiments 5-7 left the interactive framework of the previous two sets of experiments and explored spatial perspective taking in a non-interactive language task. I investigated why the presence of a person in the environment can induce speakers to abandon their self perspective to locate objects: Do speakers adapt their spatial descriptions to the vantage point of the person out of intentionality-mediated simulation or of general attention-orienting mechanisms? In an online paradigm, participants located objects in photographs that sometimes contained a person or a plant in various positions with respect to the to-be-located object. Findings were consistent with the simulated intentional accounts and linked non-self spatial perspective in language to the apprehension of another person’s visual affordance.

Experiments 8-9 investigated the role of shared experience on perspective taking in spatial language. Prior to any communicative and interactional demand, do speakers adapt their spatial descriptions to the presumed perspective of someone who is attending to the same environment at the same time as them? And is this tendency further affected by the number of co-attendees? I expanded the previous online paradigm and induced participants into thinking that someone else was doing the task at the same time as them. I found that shared experience reinforced self perspective (via shared perspective) rather than reinforcing non-self perspective (via unshared perspective). I did not find any crowd effect.
LAY SUMMARY

Nobody lives in a vacuum. Not only our perceptions, actions and thoughts but also the way we use language, i.e. how we individually choose to talk about and refer to things, is profoundly influenced by context. When we converse with someone, we tend to call a chair “a seat” if our interlocutor has done so, thus converging on each other’s choice of words (lexical convergence). We are also able to quickly process where someone else stands in space in a way that affects how we talk about that space, asking, for instance, for the cup “to the left of the teapot” if this is what it looks like from our tablemate’s perspective (non-self perspective taking in language production). This thesis is fundamentally about people (other people) as sources of contextual information and how they shape our choices of words to refer to things and talk about space.

Both lexical convergence and spatial perspective in language use have been studied from disparate perspectives and multiple explanations have been proposed for both phenomena. In this thesis, I am interested in bringing together different theoretical approaches to each and investigate what social circumstances might lead speakers to adapt linguistically to others or prevent them from doing so.

As for lexical convergence, I asked whether the pervasiveness and assumed automaticity of the phenomenon is robust against modulations in the speakers’ social context or instead is affected by them. My aim was to better understand how socially-motivated mechanisms might interplay with the proposed automatic cognitive processes for lexical convergence. I focused on contexts in which the speaker does not share the same intents as their interlocutor (but she is instead deceiving them), and contexts in which the interlocutor’s behaviour lead to a negative vs. positive impression on the speaker. My results suggests that lexical convergence is robust against modulations in the social context or, at least, speaker’s tendency to re-use their interlocutor’s lexical choices may not be affected by social factors in any straightforward way, so that non-social mechanisms may be playing a major role.

As for a speaker’s choice of perspective when describing space, there are conflicting explanations as to why people often abandon their self spatial perspective when another person is present in the environment. Building on insights from different research fields (i.e., spatial language production, spatial cognition, joint attention and joint action), I show that speakers tend to abandon their own perspective when perceiving a person that could see and potentially act on the objects, and that holds an opposite spatial perspective. I interpreted these results in terms of simulated intentional mechanisms, whereby the speaker imagine themselves in the other person’s shoes upon apprehension of this person’s visual affordance and agency. I further show that it is not necessary for the other to be physically there to affect a speaker’s choice of spatial perspective. To the contrary, my results suggest that the belief of sharing the experience with another person can reinforce taking one’s own perspective when the co-attendee also shares the speaker’s self perspective. However, it does not orient speakers towards a non-self perspective when the co-attendee holds an opposite perspective, or when there is more than one co-attendee.
ACKNOWLEDGMENTS

‘Don’t Panic’ The Hitchhiker’s Guide to the Galaxy

Completing a PhD takes more than resilience and intellectual curiosity. The support I have received throughout the past four years has been enormous and equally important.

I would like to thank my supervisors, Martin Pickering and Holly Branigan, for their guidance and continuous encouragement throughout this process. Their insights and suggestions have shaped my research and helped me become a better researcher. They have gone beyond their academic duties to fire fight my worries and concerns, and worked to instil confidence in my abilities and work. Holly, Martin, I'll “blame” you forever for turning an analytical philosopher into an experimental psychologist! :)

I am grateful to the PPLS committee that four years ago gave me the possibility to start this adventure by offering me a Principal Career's Development Scholarship. I also would like to thank the IT and the administrative staff at PPLS: they were always there to help

I am in debt to the Joint Action Reading Group, in particular Chiara Gambi, Jarek Lelonkiewicz, Alex Pap, Ed Baggs, Lucia Castillo and Assem Berniyazova for inspiring discussions (and for agreeing to embark on the crazy endeavour of science public engagement with me – I like to think we had fun). This thesis owes you all a lot.

I would like to thank Adam Moore who was a member of my first-year PhD review meeting and helped me define a new experimental paradigm after the first one did not work as expected. I am also grateful to Ian Finlayson, Moreno Coco and Martin Corley for insightful discussions about stats and helpful suggestions. And my special thanks to Felix Süßenbach who I'm sure will miss my daily statistical quibbles.

My warmest thanks go to all my flatmates and friends who have made the whole journey all the more enjoyable; in particular Andrew McFarlane, Diego Frassinelli and Eva Hasler - the most beautiful faces (and backs) I could have ever asked for experimental stimuli :P - (music) Lauren Hadley and husband (“!”; cit.) James, Roberto Grasso, Panita Suavansri (the
great tennis fan of Chapter 8), Sebastian Sandoval Simila, (monkey) Lauren Robinson, Mariana Vega-Mendoza, Iva Čukić and all the contestants in the X-ingredient cooking competition.

Il mio ultimo ringraziamento, il più importante, è per Ermanno, Flavia e Nicolò. Grazie per avermi sostenuta (e sopportata) in questi quattro anni (e i precedenti ventotto). Senza di voi, non sarei arrivata fin qui.
## CONTENTS

Declaration..............................................................................................................................................3
Abstract..................................................................................................................................................5
Lay Summary .........................................................................................................................................7
Acknowledgments ..................................................................................................................................9
Contents ..............................................................................................................................................11

### Chapter 1
Introduction...........................................................................................................................................19

### Chapter 2
Lexical and Other Linguistic Convergence: a Review .................................................................23

2.0 Chapter Overview.............................................................................................................................23
2.1 Two (plus two) families of explanations.......................................................................................23
  2.1.1 Automatic cognitive accounts.................................................................................................26
  2.1.2 The social accounts ..............................................................................................................31
  2.1.3 Audience-design accounts.....................................................................................................37
2.2 Overlap, differences and possible points of contact.................................................................40
2.3 The dynamical-system accounts..................................................................................................46
2.4 Types of linguistic repetitions.......................................................................................................48
  2.4.1 Interpersonal convergence, topicality and linguistic conventions ....................................49
  2.4.2 Short-term vs. long-term linguistic convergence ...............................................................50
2.5 The operationalisation of linguistic convergence .......................................................................51
  2.5.1 Laboratory-based operationalisations ..................................................................................51
  2.5.2 Computational operationalisations ......................................................................................54
2.6 Future directions ............................................................................................................................58
  2.6.1 Beyond cooperative truthful interactions .............................................................................58
2.6.2 Social accounts: any evidence from lexical convergence? .............................. 60

Chapter 3

Lexical Convergence in Deceptive Interaction ......................................................... 63

3.0 Chapter Overview ............................................................................................... 63
3.1 Introduction ........................................................................................................ 63
3.2 Deception and lying: an attempt at definition ..................................................... 65
3.3 Characteristics of deception and lying that may interplay with linguistic convergence mechanisms .................................................................................................................. 67
  3.3.1 A taxing affair .................................................................................................. 67
  3.3.2 Other forms of load ....................................................................................... 71
  3.3.3 Deception and lying as a form of communication ........................................ 72
3.4 Linking deception and linguistic convergence .................................................... 75
  3.4.1 Automatic-cognitive and socio-strategic accounts of linguistic convergence .... 75
3.5 Finally bringing the two together ........................................................................ 77

3.6 Experiment 1 ...................................................................................................... 80
  3.6.1 Hypotheses .................................................................................................... 81
  3.6.2 Method ......................................................................................................... 82
  3.6.3 Analysis ....................................................................................................... 89
  3.6.4 Results ........................................................................................................ 90
  3.6.5 Exploratory Analysis .................................................................................. 92
  3.6.6 Discussion ................................................................................................... 94

3.7 Experiment 2 .................................................................................................... 94
  3.7.1 Methods ...................................................................................................... 96
  3.7.2 Results ....................................................................................................... 103
  3.7.3 Discussion ................................................................................................ 108

3.8 Experiment 3 .................................................................................................... 109
  3.8.1 Methods .................................................................................................... 110
Chapter 4

Cross-Recurrence Quantification Analysis: an Introduction ..................... 127

4.1 Cross-Recurrence Plots (CRPs) and Cross-recurrence Quantification Analysis (CRQA) for discrete time series ................................................................. 127

4.1.1 (Verbal) interpersonal interaction as bivariate discrete time series .......... 127

4.1.2 Why is CRQA suitable for the analysis of conversations? .................. 128

4.2 Categorical CRQA for the study of lexical convergence: how does it work? .... 131

4.2.1 Exploring categorical CRPs and lexical coding a bit further .................. 136

4.2.2 Windowed (categorical) CRQA ....................................................... 138

4.2.3 Categorical Recurrence-Quantification Analysis (RQA) ..................... 139

4.3 Applications of CRQA to the study of interpersonal verbal interaction .......... 141

4.3.1 CRQA and conversation: the variety of non-linguistic coordination .......... 141

4.3.2 Linguistic coordination: Lab-constrained and natural conversations .......... 141

4.3.3 Statistical testing ............................................................................. 143

4.3.4 Choice of CRQA measures, effects sizes and cross-interaction variability ..... 145

4.4 Summary and conclusions ........................................................................ 146

Chapter 5

Social Impression and Lexical Convergence ................................................. 147

5.1 Theoretical Background ......................................................................... 147

5.2 Experiment 4 ......................................................................................... 153

5.2.1 Methods .......................................................................................... 153

5.2.2 Pilot (manipulation check) ............................................................... 162

5.2.3 Lexical data: Coding ....................................................................... 164
Chapter 6
Perspective Taking in Spatial Language: a Review ..........................192

6.0 Chapter Overview.................................................................192
6.1 Perspective taking in space.......................................................192
6.2 Perspective taking as a costly cognitive process: evidence from spatial cognition research.................................................................197
6.3 The simultaneous activation of multiple perspectives: evidence from spatial language processing .........................................................200
6.4 Spatial perspective taking in interactive contexts........................203
6.5 Agency, intentionality and attention-orienting cues......................206
6.6 Conclusions and future directions .............................................209

Chapter 7
Non-Self Perspective Taking in Non-interactive Language Production ..........212

7.0 Chapter Overview......................................................................212
7.1 Introduction..............................................................................212
7.1.1 The current study.................................................................216
7.2 Experiment 5...........................................................................217
7.2.1 Method...............................................................................219
| 8.2.3 | Analysis | 264 |
| 8.2.4 | Results | 268 |
| 8.2.5 | Discussion | 274 |
| **8.3** | **Experiment 9** | **275** |
| 8.3.1 | Rationale | 275 |
| 8.3.2 | Method | 276 |
| 8.3.3 | Analysis | 277 |
| 8.3.4 | Results | 278 |
| 8.3.5 | Discussion | 284 |
| **8.4** | General Discussion | 284 |
| **8.5** | Conclusions | 289 |

**Chapter 9**

**Conclusions** | **292** |
| 9.1 | Thesis Summary | 292 |
| 9.2 | Summary of the findings | 293 |
| 9.2.1 | Lexical convergence | 293 |
| 9.2.2 | Spatial language perspective | 296 |
| 9.3 | Limitations and directions for future research | 298 |
| 9.3.1 | Lexical convergence | 298 |
| 9.3.2 | Spatial language perspective | 301 |
| 9.4 | Conclusion | 304 |

**References** | **306**

Appendix A | 338
Appendix B | 340
Appendix C | 344
Appendix D | 349
Appendix E ..................................................................................................................350
Appendix F ..................................................................................................................351
Appendix G ..................................................................................................................354
Appendix H ..................................................................................................................355
Appendix I ..................................................................................................................356
Appendix J ..................................................................................................................358
Appendix K ..................................................................................................................360
Appendix L ..................................................................................................................361
Appendix M ..................................................................................................................362
Appendix N ..................................................................................................................363
Appendix O ..................................................................................................................364
Appendix P ..................................................................................................................367
Appendix Q ..................................................................................................................369
Appendix R ..................................................................................................................370
Appendix S ..................................................................................................................373
Appendix T ..................................................................................................................375
Chapter 1

INTRODUCTION

Nobody lives in a vacuum. Not only our perceptions, actions and thoughts but also the way we use language is profoundly influenced by context: what is immediately available around us or we have formerly experienced. Contextual information comes in a variety of types. It is the objects scattered around a room, the current time, your plans for the day, the curious man on the street, the conversation you had yesterday with your friend (and that still makes you laugh), how tired you feel today. This thesis is fundamentally about people (other people) as sources of contextual information and how they shape our choices of words to refer to things and talk about space. Specifically, I ask under what circumstance and how speakers are influenced by others when choosing the wording for their utterances.

The human brain is highly sensitive to social information (Allison, Puce, & McCarthy, 2000; Dimberg, Thunberg, & Elmehed, 2000) and so is our language production system: individuals adjust not just what they say but also how they say it in response to the social context. For instance, speakers quickly recognise if there are any discrepancies in the information available to themselves and their conversational partner, and adjust their communication accordingly. Speakers provide additional details to identify entities that they think are unfamiliar to their interlocutor (Fussell & Krauss, 1992; Isaacs & Clark, 1987); their choice of words (e.g., adjectives) depends on how they and their interlocutors have previously referred to an object (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986) and are shaped by their partner’s perceived language competences (Branigan, Pickering, Pearson, Mclean, & Brown, 2011; Genesee, Nicoladis, & Paradis, 1995), what their partner can and cannot see (Horton & Keysar, 1996), and the speaker’s own pragmatic goals (Yoon & Brown-Schmidt, 2013). Speakers also quickly process where someone else stands in space in a way that affects how they talk about that space, whether this someone is their addressee (e.g., Schober, 1993, 1995) or just a bystander in the environment (Tversky & Hard, 2009). So it seems that how we individually choose to talk about and refer to things is highly
sensitive to the presence of others, our conversational histories, and our interactional
expectations and goals.

In this thesis I investigate the social factors that might lead speakers to adapt linguistically to
others. The question of linguistic adaptation is conceived and addressed at two levels: with
respect to lexical convergence (i.e., interlocutors coordinating their lexical choices with each
other), and with respect to spatial perspective taking in language use (i.e., speakers
abandoning their self perspective in favour of another's when verbally locating objects in
space).

Both areas have been studied from disparate perspectives and multiple explanations have
been proposed for both phenomena. In this thesis, I am interested in bringing together
different theoretical approaches to each. Thus, what motivates my work is two-fold. First, I
aim to contribute to the understanding of the interplay between accounts of linguistic
convergence that emphasise automatic cognitive mechanisms and accounts that emphasise
socially-motivated mechanisms. In doing so, I want to explore new analytical tools
(borrowed from dynamical systems) for the investigation of interpersonal coordination in
conversation (cross-recurrence quantification analysis (CRQA)). Second, there are
conflicting explanations as to why people often abandon their self spatial perspective when
another person is present in the environment. I aim to clarify this by bringing together
insights from different research fields: spatial language production, spatial cognition, joint
attention and joint action. Experimentally, this thesis investigates the following four
questions.

Do deceiving interlocutors show a different propensity to lexically converge than truthful
interlocutors? (Chapter 3) The investigation of linguistic convergence in non-truthful non-
cooperative interactions has received little theoretical and empirical attention in the
psycholinguistics of dialogue. Here I attempt to fill this gap and compare the degree of
lexical convergence between deceptive and truthful interactions. In two novel interactive
priming paradigms, half of the participants deceived their naïve partner in a detective game
or a picture naming/matching task in order to jeopardise their partner's performance in
resolving the crime or in a related memory task. I use the case of deceptive interaction to
further our understanding of the interplay between the automatic-cognitive mechanisms and social processes that have been proposed to underlie linguistic convergence.

Does the social impression that interlocutors form of each other affect the tendency for them to lexically converge with each other? (Chapters 4-5) For this work I focus on the social accounts of linguistic convergence and ask whether speakers' tendency to match their interlocutors' lexical choices depend on the social impression that they formed of each other in a previous interaction. I also explore whether this tendency is further modulated by the nature of their interactional goal. I develop a novel two-stage paradigm in which pairs of participants first experience a collectivist or an individualistic co-player in an economic decision game (in reality, a pre-set computer programme) and then engage in a discussion of a survival scenario (this time with the real other). The first part of the survival task is open-ended (i.e., assessing the usefulness for survival of a set of items), whereas the second part drives participants towards a precise joint goal (i.e., choosing a subset of items). Through this work, I also aim to explore the potential and discuss the limitations of CRQA for the investigation of lexical convergence in laboratory-based conversation.

Does the presence of a person trigger spatial descriptions from a non-self perspective out of intentionality-mediated simulation or general attention-orienting cues? (Chapter 7) This work leaves the interactive framework of the previous two studies and explores spatial perspective taking in a non-interactive language task. Bringing together findings from spatial language production and spatial cognition, I investigate why the presence of a person in the environment can induce speakers to abandon their self perspective to locate objects. In an online paradigm, participants locate objects in photographs that sometimes contain a person or a plant in various positions with respect to the to-be-located object. The question I ask is whether speakers adapt their spatial descriptions to the vantage point of the person out of intentionality-mediated simulation or of general attention-orienting mechanisms.

Prior to any communicative demands, does shared experience affect speakers’ spatial language perspective? (Chapter 8) In this work I link together insights from spatial language production, joint attention and joint action research and explore the role of shared experience on perspective taking in spatial language. I expand the previous online paradigm and induce participants into thinking that someone else is doing the task at the same time as them. I
investigate whether, prior to any communicative and interactional demand, speakers adapt their spatial descriptions to the presumed perspective of this other person, and whether this tendency is further affected by increasing the number of co-experiencers.

I conclude this thesis (Chapter 9) with a summary of my findings, a discussion of the limitations of each study and an outline of possible future work.

A general note on the type of analyses used in this thesis: all the analyses are frequentist and are based on the estimation of a p-value for the relevant effect. A p-value represents the probability of some data given the null-hypothesis (i.e., our predictor of interest has no effect on the dependent variable). In Chapter 5 and Chapter 7, we decided to explore Bayesian Statistics and additionally introduce the use of the Bayes Factor. Unlike p-values, the Bayes Factor allows to evaluate the evidence in favour of a hypothesis (the null or the research hypothesis; Kass & Reftery, 1995). We used it in Chapter 5 and 7 to explore its use in two situations where it has been argued to provide a better alternative to frequentist significance testing. In Chapter 5, we used the Bayes Factor when the frequentist analyses led to p-values at the edge of conventionally accepted significance (0.02 < p < .05), as Bayes Factors have been shown to be more conservative than p-values (Defazio, 2016, September 13). In Chapter 7, we used the Bayes Factor when our conclusions were based on non-significant effects (e.g., lack of significant interactions); the Bayes Factor was able to confirm that the null-hypothesis could not be rejected because the null-hypothesis was indeed favoured by the data and not because the sample size was too small (Kass & Raftery, 1995).
Chapter 2

**LEXICAL AND OTHER LINGUISTIC CONVERGENCE: A REVIEW**

**2.0 CHAPTER OVERVIEW**

In this chapter I lay the theoretical groundwork for discussing the experimental work on lexical convergence presented in Chapter 3-5, and provide an introduction to the lexical and linguistic convergence literature. In 2.1 I present the various accounts of linguistic convergence that have been proposed in psycholinguistics, social cognition and cognitive science, and discuss some of the empirical evidence used to support them. In 2.2 I attempt to identify relevant overlaps, differences and points of contact between these theoretical approaches. In 2.4 I discuss some important distinctions for the study of linguistic convergence: the one between communicative interpersonal convergence and non-interactional inter-speaker repetitions; and the one between short-term and long-term convergence.

Section 2.5 describes how linguistic convergence has variously been operationalised in experimental and corpora studies, and discusses what properties these operationalisations ascribe to linguistic convergence and the underlying cognitive mechanisms. In 2.6 I outline directions for future research and discuss some of the evidence that motivated my work.

**2.1 TWO (PLUS TWO) FAMILIES OF EXPLANATIONS**

Whenever we are about to speak, there are choices to be made (Ferreira & Dell, 2000). One important choice concerns which words to use to phrase our message. Many words have meanings that are shared by other words (e.g., “cup” and “mug”) and the same object, person or idea can be denoted by a variety of potential labels (e.g., the dog ate or chewed on my thesis; Furnas, Landauer, Gomez, & Dumais, 1987), which suggests that there is potential for tremendous variability in a speaker’s lexical choices (Brennan, 1996). How do we choose? Our focus is on speakers’ lexical choices in conversation. In the following sections we look
at the evidence that suggest that people tend to re-use words that their interlocutor has just used; a phenomenon called *lexical convergence*.

A terminological note before we proceed: throughout this thesis we will refer to the observed behaviour of two interlocutors’ matching each other’s lexical (or other linguistic) choices as *convergence*. We use it as a theoretically neutral term that does not endorse any specific underlying mechanisms among the ones proposed. Specifically, we reserve the term *alignment* to denote the alignment of linguistic representations and situation models as proposed within the Interactive Alignment Model (Pickering & Garrod, 2004; see section 2.1.1).

While the degree of lexical repetitions is low *between* conversations (even conversations about the same topic), it is much higher *within* conversations (e.g., Brennan & Clark, 1996; Garrod & Doherty, 1994). Conversing makes individuals’ language use similar. For instance, interlocutors tend to develop common description schemes to denote complex visual stimuli (e.g., mazes, tangram pictures). They were found to converge on both using "box" or both using "square" to refer to a maze’s nodes (e.g., Garrod & Anderson, 1987; Garrod & Clark, 1993; Garrod & Doherty, 1994) and on calling a tangram picture "the bat" rather than "the anchor" or "the Olympic torch" (e.g., Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). Evidence is not limited to referring expressions. Interlocutors were found to converge on expressions designating confidence (e.g., using phrases containing "sure" rather than "convinced"; Fusaroli et al., 2012). Lexical adaptation does not only happen over the course of a whole conversation but also turn by turn, with the use of a certain word by one interlocutor triggering the use of the same word by the other in the immediately following utterances. Asked “*At what time do you close*?”, Dutch vendors provided congruent answers (e.g., “*At five o’clock*” rather than “*Five o’clock*”; Levelt & Kelter, 1982). And speakers tended to call a handled cylindrical container a “*mug*” rather than a “*cup*” if their dialogue partner had done so in a previous turn (e.g., Branigan et al., 2011; see also Brennan & Clark, 1996).

That interlocutors tend to re-use each other’s words has long been seen as part of a broader tendency that people in conversation, by the pure fact of communicating with each other, become similar at a number of para-linguistic (e.g., accent, pitch, speech rate, loudness) and
linguistic (e.g., word choice, syntax) levels. Interlocutors do not only pick up and reuse each other's lexical choices (Brennan & Clark, 1996) but also syntactic structures (Branigan, Pickering, & Cleland, 2000) and ways of conceptualising the environment (Garrod & Anderson, 1987). Their speech rate (Giles, Coupland, & Coupland, 1991), pitch (Natale, 1975) and accent (Bourhis & Giles, 1997) become more similar. Why is this happening?

In this thesis we are concerned with two of the proposed explanations for lexical, and more generally linguistic, convergence. The first is that our language processing system is hardwired to pick up and reuse recently processed linguistic input and so converging comes naturally, effortlessly and (mostly) without control (automatic cognitive accounts; e.g., Pickering & Garrod, 2004). The second is that we converge because converging makes us socially more attractive and so, by converging, we can better navigate the social world (social accounts; e.g., Giles & Coupland, 1991). There are two other families of explanations to lexical and linguistic convergence that won't be directly tested in the experimental work of this thesis but that will come out in the discussion of our findings. According to these models, lexical convergence is either an outgrowth of conscious strategies (grounding) to enhance the effectiveness of communication (audience-design accounts; e.g., Brennan & Clark, 1996), or part of a bigger set of coordinative dynamics, not all of which involve “doing the same thing”, which self-arise when individuals interact (dynamical-system accounts; e.g., Fusaroli, Raczaszek-Leonardi, & Tylén, 2013).

It should be clear already from the previous paragraph that people believe that there is more than one mechanism as to why speakers converge on each other's lexical choices, and, more generally, on each other's linguistic style in conversation. Despite the claimed differences, all these accounts seem to interpret lexical (and more generally linguistic) convergence as a healthy characteristic of the conversation between two speakers. However, most of the supporting findings to these different accounts come from studies that have assumed the truthfulness of the speakers in conversations, and had them engage in cooperative tasks that promoted a positive or at worst neutral disposition toward each other. It is thus an open question whether a similar degree of lexical (and more generally linguistic) convergence would be observed in contexts that disrupt this underlying assumption of cooperation and that do not encourage positive rapport between the interlocutors. This issue is addressed in
Chapter 3 and Chapter 5 (see also section 2.6.1), where we investigate whether conversations in which one interlocutor is deceiving the other or which prevent from creating rapport present similar degrees of lexical convergence as truthful cooperative conversations and positively-valued interaction.

Of course, the different mechanisms proposed do not have to be mutually exclusive. In the next sections we revise the four accounts' main claims and the evidence that supports them. The aim is to provide some conceptual clarity, and attempt to identify overlap, differences, and possible points of contact between the accounts.

In revising the evidence in favour of the different accounts, we try to prioritise the studies that addressed lexical rather than syntactic or para-linguistic convergence because of their relevance for our experiments. However, as will later be apparent, this is not always possible as evidence for lexical convergence is sometimes lacking. We discuss in section 2.6.2 whether findings about other types of convergence can be extended to convergence on lexical choices.

2.1.1 Automatic cognitive accounts

Cognitive accounts were proposed in psycholinguistics and explain lexical and other linguistic convergence in terms of automatic mechanisms grounded in the architecture of our cognitive system: the automatic re-use of transiently activated memory representations (Pickering & Garrod, 2004) or implicit learning (Bock & Griffin, 2000; Chang, Dell, & Bock, 2006).

The interactive alignment model

According to the Interactive Alignment Model (IAM; Pickering & Garrod, 2004), linguistic convergence is the result of speech production and comprehension building on shared memory representations of linguistic information, and on memory retrieval favouring recently activated material. In dialogue, both production and comprehension systems are active at once. When the partner's speech is processed, corresponding internal representations get activated at phonetic, phonological, lexical and syntactic levels; these lower-level representations then contribute to establish higher-level semantic and discourse
representations of the speech content. As in dialogue a person regularly alternates between the roles of listener and speaker, she is likely to produce speech while the linguistic representations triggered during comprehension of her partner's speech are still active in her memory: the corresponding linguistic material is thus more likely to be reused compared to alternatives that are less or not active (automatic priming mechanisms). The same priming process is at work for the individual speaker's speech production, who is thus more likely to re-use linguistic material that she has already produced herself. This priming mechanism is assumed to lead to percolation across linguistic levels: convergence (and alignment) at one level (e.g., lexical) would induce alignment at other levels (e.g., syntactic, as in “lexical boost” effects; Cleland & Pickering, 2003).

There are three effects of this priming mechanism. Externally, dialogue partners will tend to re-use previously heard and produced linguistic material and we observe convergence of their linguistic features (interpersonal convergence), as well as consistency in the individual speaker's own production (self-convergence or self-priming). Internally, interlocutors will activate the same linguistic representations (linguistic alignment), which will lead to the alignment of their situation models (i.e., a common understanding of the situation they are discussing, implicit grounding). Interpersonal priming, and the linguistic convergence and alignment that are derived from it, is the fulcrum of our capacity of successful interaction. Specifically, it benefits communication in two ways. On the one hand, it eases the processing effort of the interlocutors: when listening, it increases their chance to correctly understand what their partner means (e.g., if we have previously talked about Donald Duck as “Donald”, you are later more likely to correctly infer that I am referring to the fictional duck and not Trump when I say “Donald even got 133 write-in votes in the 2014 Swedish national elections”); when speaking, it eases their computational load as they can stick to the already established ways to refer and conceptualise things (Branigan et al., 2000). On the other hand, because linguistic convergence eventually helps interlocutors build common understanding, priming also contributes to communicative success (Pickering & Garrod, 2006). Importantly, the process is automatic and does not require any intentional control by the speakers or any
explicit consideration of their partner's communicative needs (e.g., what they would more easily understand).\(^1\)

The implicit learning model

We record that there is another purely cognitive explanation of linguistic convergence: the implicit procedural learning account (e.g., Bock & Griffin, 2000; Chang, et al., 2006; Jaeger & Snider, 2013). Here, linguistic convergence results from an error-based implicit learning mechanism. Exposure to linguistic input (i.e., an interlocutor's speech) changes the organisation of the linguistic abstract representations in the comprehender; like nodes that receive more weight in a connectionist network, the just processed linguistic procedures are favoured for future use. At the same time, this creates expectations concerning what is likely vs. unlikely linguistic input (e.g., “cup” over “mug”). If the speaker is then exposed to unlikely linguistic information given the created expectations (i.e., “mug”), they will encounter a bigger prediction error which will cause a bigger re-asset of the linguistic abstract processing, resulting in a greater bias towards that unlikely material.

This account shares with IAM the view that linguistic convergence is hardwired in our language processing system, that it is an automatic process out of the control of the speaker and that it is bi-directional (i.e., both interlocutors contribute to mutual convergence). Unlike IAM, however, the implicit learning account has so far only concerned with syntactic categories and does not account for self-priming (as you cannot erroneously predict what you are about to say). We won't address this account in this thesis.

Evidence in favour of the IAM

Probably the major source of evidence in favour of the IAM and its claims that linguistic convergence is an automatic (default) mechanism that acts as the linchpin of our interactional ability is that linguistic convergence is a pervasive phenomenon that has been

\(^1\) Explicit mentalising about the partner's state of knowledge would be invoked only when alignment failures are particularly severe or out of the speakers' initiative not to align. The view that explicit adjustments to an interlocutor's knowledge/needs (when different from the speaker’s) constitute an effortful process that speakers engaged in at a later stage as a repair tool in case of clear misunderstanding finds support in other two-stage models of language use in conversation, mostly developed for comprehension (e.g., Horton & Keysar, 1996; Keysar, Barr, & Horton, 1998). We won't address these models in this thesis.
observed across all levels of language production: from phonetic (e.g., accent; Giles, 1973) and phonological features (e.g., word pronunciation; Pardo, 2006), to grammar (Branigan et al., 2000), lexicon (word choice; Brennan & Clark, 1996), prosody (e.g., pause length; Cappella & Planalp, 1981) and semantics (e.g., description schemes; Garrod & Anderson, 1987). This pervasiveness does not only concern verbal modalities but also interactional contexts. Linguistic convergence has been observed between adults in spontaneous (e.g., Reitter & Moore, 2014) and laboratory-based (e.g. Branigan, et al., 2000) conversations; between children and adults (e.g., Dale & Spivey, 2006), and patients with recognized deficits in social cognition and communication like ASD (Hopkins, Yuill, & Keller, 2016) and Asperger (Slocombe et al., 2013). Human speakers converge on the language of dialogue systems (Bell, Gustafson, & Heldner, 2003), computers (Brennan, 1996) and human-like avatars (Heyselaar, Hagoort, & Segaert, 2015), and across a variety of channels: in spoken (Hartsuiker, Bernolet, Schoonbaert, Speybroeck, & Vanderelst, 2008) and computer-mediated written communication (e.g., Branigan et al., 2011; Pickering & Branigan, 1998), including typed-in chats (Foltz, Gaspers, Thiele, Stenneken, & Cimiano, 2015a), online forum (Wang, Yen, & Reitter, 2015) and social media exchanges (Danescu-Niculescu-Mizil, Gamon, & Dumais, 2011). Note, however, that some corpus studies have cast doubts about the magnitude of linguistic convergence (especially at syntactic level) in open-ended spontaneous dialogue (e.g., Healey, Purver, & Howes, 2014; Howes, Healey, & Purver, 2010).

What further supports the low-level automaticity of the phenomenon is that speakers tend to re-use the heard linguistic material (e.g., “mug” over “cup”) also in non-interactive or socially impoverished contexts (Weatherholtz, Campbell-Kibler, & Jaeger, 2014). For instance, speakers were observed to favour the syntactic choices previously heard in the unrelated pre-recorded utterances of a speaker (Lev-Ari, 2015). Additionally, the tendency to linguistically converge seems to develop early in humans (Garrod & Clark, 1993; Mani, Durrant, & Floccia, 2012), and at an age when children have not yet developed mentalising or other-monitoring abilities (Branigan, Tosi, & Gillespie-Smith, 2016); these are all findings that support the default nature of the effect (in line with evidence about imitative motor behaviours; e.g., Kuhl & Meltzoff, 1996). For instance, Dale and Spivey (2006) tracked the degree of syntactic convergence in three longitudinal corpora of dialogues.
between children and caregivers and found that less-developed children converged more, whereas older children’s and adults’ use of grammatical forms mirrored each other less (see also Cox & van Dijk, 2013).

Finally, there is evidence that the magnitude of linguistic convergence between interlocutors is predictive of their communicative success. Ward and Litman (2007) found that the degree of lexical convergence between students and their physics computer tutor positively predicted the students’ learning outcomes. Participants instructing each other drawing routes on a map drew more accurate paths the higher their degree of lexical and syntactic convergence (Reitter & Moore, 2007, see also Carbay & Tanenhaus, 2011), and convergence at morphemic and prosodic level positively predicted a dyad’s success in a joint decision-making perceptual task (Fusaroli & Tylén, 2016).

To summarise, in line with the automatic cognitive account, evidence suggests that lexical (and linguistic) convergence is a default phenomenon that is pervasive in conversation. By reusing words that their interlocutor has already produced, because they are cognitively more available, speakers are more likely to produce messages that their interlocutor can understand without designing them explicitly for them. Talking similarly leads interlocutors to think similarly, and so to communicate successfully (Menenti, Pickering, & Garrod, 2012).

The IAM’s claims have not gone unchallenged. One main observation that has been made is that despite the extensive evidence for the pervasiveness and automaticity of the phenomenon, some cases of inter-speaker linguistic convergence appear to be driven by partner-specific considerations (linked to socio-affective or communicative goals) that may not be fully captured by the automatic, effortless and default process described by IAM. In Chapter 3 and Chapter 5 we ask whether the pervasiveness and assumed automaticity of the phenomenon is robust against modulations in the speakers’ social context or instead is modulated by them. Specifically, we focus on contexts in which the speaker does not share the same interactive intents as their interlocutor (but she is instead deceiving them), and contexts in which the interlocutor’s behaviour lead to a negative vs. positive impression on the speaker. Experimental evidence is scarce in relation to both issues. But first, in the next section we address the partner-specificity concerns raised so far in the literature and the
theoretical accounts that support them. For reasons of space, we give more emphasis to the social accounts given its centrality in the empirical work of this thesis.

2.1.2 The social accounts

Accounts in social psychology and sociolinguistics have interpreted linguistic convergence as part of an individual's broader tendency to match their interacting partner's behaviour that is linked to socio-affective dynamics. The social-accounts literature is prolific and complex: different authors have provided different explanations, seemingly contradictory. In this section, we make the distinctions that we think are more relevant. Importantly, it is not our main interest to distinguish these different explanations in our empirical work.

According to a part of the social-accounts literature, which we will refer to as the proactive route (e.g., the Communicative Accommodation Theory (CAT); Giles, 1973; Giles, Mulak, Bradac, & Johnson, 1987), speakers use verbal convergence as a tool to meet social psychological outcomes and manage social relations. Linguistic convergence is explained in terms of social mechanisms that act largely separately from the possible automatic processes reviewed earlier (i.e., priming and implicit learning). Accordingly, speakers may copy key features of their interlocutors' verbal style (e.g., re-using their choice of words) as a way to pursue affective goals (Giles & Smith, 1979); for instance, to win trust, create rapport or foster affiliation as a result of appearing more similar to the interlocutor (Coupland, Coupland, Giles, & Henwood, 1988). Speakers would exploit the fact that people tend to prefer others whom they perceive more similar to themselves (e.g., Giles & Smith, 1979), and that speakers who show higher degree of similarity by matching their interlocutor's linguistic style are indeed perceived more positively (e.g., Bradac, Mulac, & House, 1988; Ireland et al., 2011; Welkowitz & Feldstein, 1970). The magnitude of the convergence is assumed to be proportional to the need of gaining the other's social approval (Giles & Coupland, 1991, Larsen, Martin, & Giles, 1977). Similarly, speakers may use linguistic convergence as a means to mark affiliation or signal identification (especially at cultural/group-membership level) with their interlocutor by sounding more similar to them (Giles & Powesland, 1975).
Another part of evidence in the social-accounts literature suggests that linguistic (and behavioural) convergence is an automatic process which is though modulated by socio-affective dynamics (modulatory route). At a theoretical level, this literature has mostly addressed behavioural (e.g., unintentional motor movements) convergence but the accounts can be extended to linguistic behaviour as well (as the reviewed evidence shows, see below). A person’s tendency to match another’s behaviour is assumed to be a natural consequence of the fact that perceiving a behaviour triggers the automatic activation of the same behavioural response (Bargh & Chartrand, 1999; Dijksterhuis & Bargh, 2001). Social factors would modulate the degree of this automatic tendency by strengthening the perception-production link in either of two ways. According to some authors (e.g., Dijksterhuis & Bargh, 2001), the social modulation is goal-independent and mechanistic: when we like or perceive someone positively, we pay more attention to what we perceive of their behaviour. The increased attention would lead to a stronger activation of the perceptual representations (e.g., of the incoming speech) and therefore to a stronger (convergent) behavioural response (given that perceptual and production representations overlap). According to other authors (e.g., Lakin & Chartrand, 2003), the social modulation is goal-driven: when we perceive someone positively, we may wish to minimise social distance or express affiliation; these implicit social goals would reinforce the perception-production links and increase behavioural (and linguistic) convergence.

To summarise the main points: for the proactive route, linguistic convergence is a tool whose use is triggered by social goals (e.g., bonding with others), so that the social component is the driving force of behavioural matching. For the modulatory route, linguistic convergence is in essence an automatic process but whose strength is moderated by the social context.

Two important points before we move on. First, even when linguistic convergence is assumed to be goal-directed (e.g., proactive route and one interpretation of the modulatory route), the social accounts do not assume it to be conscious and/or deliberate. Goal-directed processes do not necessarily entail conscious deliberate decisions. For linguistic convergence, deliberation would mean that speakers consciously focus on their partner's linguistic production and equally consciously adjust features of their own production to match it. Authors have argued that even if the goal might be conscious (e.g., being liked by
others), the strategy used (e.g., linguistic convergence) can be fully unconscious (Giles et al., 1991); others have proposed that both social goals and their pursuits can operate non-consciously (e.g., Bargh, 1994; Custers & Aarts, 2010; Lakin & Chartrand, 2003; Nisbett & Wilson, 1977). Second, both social mechanisms concede that speakers may not only converge more when perceiving or wishing affiliation, but may equally converge less, or not at all, when perceiving or wishing disfavour (see evidence in next section). Specifically, the proactive-route literature recognises that speakers may diverge from their interlocutors’ linguistic style (i.e., maximise the differences) as a form of dissociative tactic: when they wish to dissociate or mark disaffiliation or intergroup vs. outer-group dynamics (Giles & Powelsand, 1975; see also Bourhis, Giles, Leyens, & Tajfel, 1979).

We leave to section 2.2 our discussion of how the social accounts may relate to the automatic cognitive models (i.e., IAM) and the audience-design accounts. In the next section we present the evidence that supports the two social routes to linguistic convergence.

**Evidence in favour of the social accounts**

Most of the evidence in support of the social accounts of verbal convergence regards what we may call contextual social variables: characteristics of the situation in which the conversation unfolds that are linked to the way speakers perceive and manage interpersonal relationships (e.g., the interactional goal of the conversation and speakers’ impression of each other). In reviewing the findings in favour of these two social claims, it will be apparent that evidence rests mostly on para-linguistic aspects of speech (e.g., accent, pitch) and, only more recently, syntactic choices. There is no direct evidence at the lexical level, at least in terms of coordination of word choices. Furthermore, most findings are derived from correlational studies that have difficulty elucidating the directionality of the affiliation-convergence link.

**The modulatory route**

That speakers converge more on their partner’s verbal behaviour when they perceive them positively (e.g., they form a good impressions of them) and converge less (or diverge) when they perceive them negatively finds extensive support at acoustic-phonetic level (e.g., accentedness, vowel pronunciation, speaking rate, F0, intensity; see Pardo, 2013 for a full list of recent investigations). For instance, Welsh speakers who listened to the recording of
an English speaker (with whom they never interacted) converged less with his accent if he expressed bias against their culture compared to when he expressed a neutral stance, probably as a way to express disaffiliation (Bourhis & Giles, 1977; see also Doise, Sinclair, & Bourhis, 1976). Pardo, Gibbons, Suppes, and Krauss (2012) observed that college flatmates, originally unacquainted, showed small but consistent convergence to each other's pronunciation after one semester of cohabitation, an effect that was stronger the closer they felt. In a word shadowing task, the extent with which female speakers converged to the vowel pronunciation of a male talker was modulated by how attractive they found him (Babel, 2012); and in a priming experiment, participants converged to the unusually long voice-onset-times of a narrator toward whom they developed a positive disposition, a factor that reflected the combination of the intrinsic characteristics of the narrator (e.g., sexual orientation), the content of the narrative (e.g., whether the narrator behaved nicely or unpleasantly) and the attitude of the participants themselves even if in the absence of expectation of interaction (Abrego-Collier, Grove, Sonderegger, & Yu, 2011).

Turning to more content-related linguistic aspects, Moscoso Del Prado Martin and Du Bois (2015) found that the extent to which interlocutors' words expressed similar high levels of positive affect was correlated with their degree of syntactic convergence. Importantly, the effect was independent of convergence at a lexical level (i.e., it was not a consequence of speakers simply producing the same words). At a lexical level, Gonzales, Hancock, and Pennebaker (2010) observed that speakers who had a similar rate of use of functional words felt more cohesive as a group in a joint information search task: they liked each other more, and more effectively interacted and shared information (but note that this method only provides a coarse measure of lexical convergence, see section 2.5.1).

Taken together, these last studies suggest that degrees of spontaneous coordination in linguistic behaviour can reflect underlying social dynamics. However, the evidence was mostly correlational and was not clear on the directionality of the effect: did positive affect lead people to converge more as a response (as the modulatory routes would predict), or did people who happened to converge more (for any other reasons, such as perhaps implicit learning or automatic priming mechanisms) end up developing positive rapport?
More recent evidence tried to make the directionality of the effect clearer and investigated whether when a person is perceived positively, speakers may tend to decrease linguistic (and behavioural) distance by converging more with the person. Studies experimentally manipulated the extent to which speakers may develop positive affect for their interlocutor, and then measured whether the developed interpersonal bond predicted their degree of linguistic convergence in a subsequent task. However, findings have no easy interpretation and offer some puzzling results. Weatherholtz et al. (2014) had participants listen to the recordings of an ideologically-charged narrator who produced only one of two alternative syntactic structures (preferred vs. dispreferred, between-subjects). Interpersonal similarity was based mostly on political ideological agreement. Participants who heard the narrator using dispreferred syntactic structures converged less on her syntactic choices the more similar they felt to her, while there was no effect of degree of similarity on convergence on the preferred structures. In contrast, the smarter the participants judged the narrator, the less they converged on her use of preferred syntactic structures but there was no effect of perceived smartness on convergence on dispreferred structures. Furthermore, effects were overall very small. In a similar two-stage non-interactive paradigm, Lev-Ari (2015) presented participants with the recording of a (fictitious) previous participant who was presented as either smart or not so smart. The recordings primed participants with one of two syntactic alternatives. Participants then performed a written scrambled sentence task, in which syntactic convergence was measured. The more participants reported liking the speaker, the more they converged on his syntactic choices. However, the effect was restricted to the not-so-smart speaker, arguably because participants were able to empathise with him.

**The proactive route**

Evidence that verbal convergence may be proactively used by speakers as a means to pursue social goals and manage social relations (and it is not just an automatic response whose degree is modulated, e.g. by the perceived attitudes of others) concerns mostly acoustic-phonetic features and focuses on how the perception of interlocutors' relative social status or dominance in interaction could affect who is converging with whom and the magnitude of this convergence (Bourhis, Giles, Lambert, 1975; Pardo, 2006, 2013). Results suggest that it
is usually the speaker with a perceived lower status who converges toward the speaking style of the higher-status or more dominant interlocutor (Giles, 1973). For instance, Gregory and Webster (1996) found that guests at a TV show tended to convergence on the pitch of higher-status but not lower-status speakers, and bilinguals in industrial settings converged more to the first language of those who were their superiors but not their subordinate (Giles, Coupland, & Coupland, 1991). Interviewees, who are likely to desire to be liked, converged to the intensity of pronunciation of interviewers but not the other way around (Natale, 1975; see also Putnam & Street, 1984). Likewise, speakers may verbally diverge from their interlocutor to signal disaffiliation or mark disassociation. This was observed when speakers interacted with out-group members in ethnically salient contexts (Bourhis, Giles, Leyens, & Tajfel, 1979), and (at a nonverbal level) in interactions between adolescents and their parents (presumably to mark personal identity and independence; Giles & Wadleigh, 2008).

The evidence concerning linguistic aspects of speech (i.e., grammar and lexicon) is scantier and less consistent. Again, the observational nature of most studies can't elucidate the directionality of the effect. Coyle and Kaschak (2012) found that male (but not female) speakers converged less with the unusual syntactic choices of a female confederate if the confederate was in her fertile period. This was taken as a sign that males unconsciously yet strategically diverged from their partner's linguistic style to show potential positive mate characteristics (i.e., non-conformity behaviour). Pragmatic convergence was employed by members of an online health forum when providing support and showing empathy (Wang et al., 2015): responders aligned on the type of support (emotional versus informational) given by earlier respondents. Interestingly, emotional (but not informational) support was associated with higher degree of lexical (but not syntactic) convergence: respondents re-used each other's words when providing support, probably in an attempt to show bonds with each other. Note that a recent study that tried to determine the directionality of the effect and explicitly gave speakers the goal to enhance rapport in their interlocutor in an interview setting (Schoot, Heyselaar, Hagoort, & Segaert, 2016) failed to find any effect of speakers' desire to be liked on their degree of syntactic convergence.
Finally, separate evidence indicates that the link between convergence and socio-affective dynamics may not be as straightforward as the proactive and modulatory routes might suggest: stronger convergence may not consistently relate to positive dynamics and weaker convergence (or absence thereof) may not consistently relate to negative dynamics. There is space for a third route, which we might call the compensatory strategic route: speakers may use linguistic convergence as a tool to (re-) establish smooth interaction, in particular when addressing un-cooperative or unfriendly interlocutors. For instance, speakers converged to a greater extent with the lexical choices of an interlocutor that showed annoyance and impatience compared to friendliness and understanding, probably using convergence to ensure a trouble-free interaction with what was perceived as a difficult interlocutor (Balcetis & Dale, 2005).

To conclude, as already highlighted in various points throughout this section, evidence in favour of these social interpretations of linguistic convergence is limited to sound-related aspects of speech (e.g., accent and pronunciation) and, more recently, syntactic features. Evidence at the lexical level has only so far concerned speakers’ rate of use of lexical categories; a coarse characterisation of speakers’ lexical production that does not capture the structures in the repetition of the lexical choices between speakers and their evolution over time. Furthermore, as already noted, the correlational nature of most these studies prevented identification of whether it was indeed the nature of the rapport between the dialogue partners that determined their degree of linguistic convergence (or it was instead the other way around). Both these issues are addressed in Chapter 5, where we first manipulate the speakers' social perception of their dialogue partner and then measure the degree with which they converge with their partner's lexical choices. But first, in the next section, we discuss a different set of accounts that link linguistic convergence to the pursuit of goals, this time of a communicative nature.

2.1.3 Audience-design accounts

Audience design accounts, like Brennan and Clark's (1996) conceptual pact and Clark's (1983; 1996) optimal design models assume that dialogue requires a process of negotiation
between interlocutors which has the convergence of linguistic expressions as a side outcome. According to these models, beliefs about the dialogue partner's state of knowledge (mentalising) provides immediate and constant guidance to online language processing, both in production and comprehension. Thus, the choice of a speaker to re-use the same linguistic expression previously used by their interlocutor would be mediated by explicit considerations of communicative design: recognising that the other may have a different understanding of the situation, modelling this knowledge as different (or shared), and adjusting their message accordingly to facilitate communication. Through this constant process of other/self-monitoring and (mostly tacit) negotiation (grounding), partners achieve a joint perspective on the object of their communication. Through the process of grounding, dialogue partners converge at various linguistic and paralinguistic levels (i.e., repetitions are a sign of achieved grounding; Brennan, Galati, & Kuhlen 2010). Importantly, audience-design accounts assume that linguistic convergence is partner-specific and results from a voluntary and controlled strategy to foster mutual understanding.

**Evidence in favour of audience-design accounts**

Support for the audience-design accounts comes from two sources of evidence. In line with the idea that interacting speakers form conceptual pacts with each other (and not with someone else) (Clark & Schaefer, 1989; Schober & Clark, 1989), studies have found that how a speaker refers to objects is affected by the conversation history she shares with a particular interlocutor. Experimental studies usually have pairs of participants do a task together which requires taking turns to refer to objects over several trials (e.g., describing and matching pictures of real or abstract objects). In Brennan and Clark (1996), participants described pictures to each other in order to arrange them in the same order. In a first set of trials, speakers could use basic-level category terms to point out objects (e.g., the shoe to refer to the only shoe among other objects). In a subsequent set of trials, different objects of the same category were present (e.g., different types of shoes). Speakers adjusted by adding additional information to their descriptions (e.g., the high-heel shoe). After this, they went on

---

2 Note that the grounding process does not have to be explicit (with speakers overtly signaling that they have understood what the partner is referring to and that they will be using the same word according to the same acceptation); interlocutors may tacitly agree on how to refer to things (cf. Brannan, Galati, & Kuhlen, 2010).
playing with either the same interlocutor or a different one in a context as in the first set of trials where basic-level category terms were sufficient to correctly identify objects. Speakers showed conversation (or partner) consistency: they kept using the over-specific terms ("high-heel shoe") with the old partner, but often reverted to the basic-level category terms with the novel partner.

In line with the claim that utterances are planned with a specific interlocutor in mind (Clark, 1996), studies have found that the tendency of speakers to adapt to the linguistic choices of their interlocutors is mediated by the impressions that they gain of the other as a speaker. This is in line with a broader set of evidence that speakers formulate utterances taking into account what they believe their interlocutor may or may not know (e.g., Fussell & Krauss, 1992; Isaacs & Clark, 1987; yet see Horton & Gerrig, 2005, for a conflicting position). For instance, an interactive priming study in which participants took turn with a partner at naming pictures for each other found that the extent with which participants converged on the lexical choices of their partner was proportional to how poor they judged their partner's communicative capacity (Branigan et al., 2011): participants converged more with computers than human partners, and even more with less advanced computers. Bortfeld and Brennan (1997) observed that native speakers abandoned idiomacity to converge on the more descriptive denotations used by their non-native interlocutor (e.g., abandoning rocking chair in favour of the chair that goes back and forth), presumably to guarantee their partner's understanding (see also Ferguson, 1975).

It must be noted that others have suggested that partner-specific lexical adaptation can be explained in terms of general memory processing. Speakers would create automatic memory associations between a certain context, a certain linguistic expression and a certain interlocutor, and simply remember how a certain interlocutor had previously named a given object (Horton & Gerrig, 2005a, 2005b). However, recent studies have found equally strong convergence effects in a syntactic priming task when prime and target pictures were different tokens of the same object type (e.g., pictures of two different cups) as when they were the same token (e.g., the same picture of the same cup; Branigan et al., 2016). It is hard to

---

3 Note, however, that these adjustments may have been prompted by the feedback received from the new addressee (Barr & Keysar, 2006; Horton & Gerrig, 2002).
combine these results with the memory retrieval cue accounts as these accounts would predict stronger convergence effects when naming the same token (as the word-picture-person compound cue should be stronger; see also Branigan, Pickering, McLean, & Cleland, 2007).

2.2 **OVERLAP, DIFFERENCES AND POSSIBLE POINTS OF CONTACT**

In the previous sections, we discussed three of the main explanatory frameworks of lexical and linguistic convergence. Those considerations will become relevant in Chapter 3 and Chapter 5 where we investigate how automatic and socially-mediated mechanisms may interfere with lexical convergence when modulations in the speakers’ social context occur. Here, we attempt to identify overlaps, differences and points of contact between them. We leave the fourth (dynamical-systems) explanation to the end as it may provide some answers to the open questions that will arise in this section.

A recap of the different accounts’ main claims is probably due. For the IAM, linguistic convergence is a default behaviour driven by purely automatic mechanisms (i.e., priming). Linguistic convergence leads to and simultaneously is a sign of alignment of mental representations: interlocutors who speak similarly will think similarly. The proactive route within the social accounts understands linguistic convergence as a goal-driven behaviour: speakers exploit the fact that people like others similar to themselves and so copy their interlocutors’ linguistic style to ingratiate with them. The modulatory social routes interpret linguistic convergence as a spontaneous mechanism (rooted in perception-production links) whose magnitude is though modulated by the perceived quality of the social context and/or affiliative goals; social factors modulate the degree of convergence but the mechanism does not require a motivation to affiliate or social goal or perceived positive affect to occur. Audience-design accounts see convergent linguistic behaviour as partner-specific adaptations that are not automatic but are brought about by explicit considerations of what the other is more likely to understand or what has been said before with that person.

Under the IAM, linguistic convergence is a symmetrical process that takes place in both interlocutors at the same time. Of course, different situations may make the convergence effect larger one way (e.g., you are talking with your mother while she is trying to solve her
daily sudoku and so she is not paying too much attention) but the model assumes that it should occur both ways; moreover, this is true even at a moment by moment level because comprehension and production both enhance activation. Without this symmetry, in fact, there would be no alignment of linguistic and discourse representations (if not by chance), and mutual understanding would be at risk.

In contrast, audience-design and the proactive social mechanisms share the view that linguistic convergence is, in essence, a goal-driven asymmetrical process (cf. Krauss & Pardo, 2004), whereby one interlocutor (independently of the other) adjusts linguistically to the other guided by her own relational (proactive social route) or communicative goals (audience-design accounts). Of course, if driven by the same communicative or affiliative goals (e.g., accomplish a joint task), interlocutors could “meet in the middle” and adjust both ways, but in principle these mechanisms may involve convergence for one interlocutor but not the other. For the modulatory social mechanism, some degree of linguistic convergence should occur in both interlocutors (via the automatic link) but the strength of the convergence is an asymmetrical process (i.e., it is a person’s individual response to how she personally perceives her social context) which can develop to different extent in the two speakers.

These are important distinctions as they give different weight to a speaker's expectation that the other will perceive their convergence under the automatic compared to both the social proactive and audience-design framework (see below for the modulatory route). According to the social-proactive mechanism, linguistic convergence is a signal sent by speakers when they are aware that it may be received (Toma, 2014): only if received, speakers can meet their goal (e.g., be liked) Similarly, Bavelas, Black, Chovil, Lemery, and Mullett (1988) have argued that the person whose behaviour is being matched needs to be able to notice these converging behaviours and perceive them as markers of the positive development of the interaction. And in fact, work under the social (proactive) frameworks has concentrated on measuring the perception of convergence and its consequences on the perceiver’s impressions about their interlocutor and the interaction (e.g., Pardo, 2013; Giles et al., 1987). The audience design accounts also see convergent linguistic behaviour as a communicative signal (e.g., of shared meaning/understanding) that is supposed to be discerned as such by
the interlocutor (Krauss & Pardo, 2004). In contrast, because convergent linguistic behaviour arises automatically via priming under the IAM, convergence is not restricted to situations where it can be perceived. This claim finds support in the many studies that found convergence effects (e.g., Namy, Nygaard, & Sauerteig, 2002; Nielsen, 2011; Shockley, Sabadini, & Fowler, 2004) and effects of social factors on linguistic convergence (e.g., Bourhis & Giles, 1977; Lev-Ari, 2015) in contexts where the speakers did not expect any interaction and there was no recipient for their convergence (i.e., they passively listened to voice recordings). It is harder to account for these findings within the social proactive or the audience-design framework. However, these findings could still be consistent with the modulatory social mechanism; here social factors are interpreted as possible sources of modulation that act on top of underlying automatic converging mechanisms (cf. Weatherholz et al., 2014).

The three frameworks also assign different roles to divergence and, in general, to cases when speakers fail to adjust to the linguistic style of their interlocutor. Let us start by noticing that for discrete linguistic behaviour like syntactic and lexical choices, on a single occasion, a speaker can converge or fail to converge with their interlocutor’s choice (e.g., if I ask you to “pass me the mug”, you can answer “which mug” or “which cup?”). Over a conversation, convergence can be better represented as a continuum, with different conversations showing variation in convergence rate. On some occasions interlocutors may show reduced or increased convergence; at the extreme they could maximise the linguistic difference with their partner (divergence). Note that it is hard to establish empirically what constitutes reduced or increased convergence⁴, at least without knowing in advance speakers’ production baseline/preference or comparing different situations.

According to the social accounts, failing to converge and reduced convergence (and divergence) serve a clear communicative purpose (i.e., to express and achieve social distance, e.g., Bilous & Krauss, 1988; Bourhis, Giles, Lambert, 1991). In addition, modulations in the rate of linguistic convergence are a key aspect of the modulatory social mechanism that relate to the valence of the perceived social environment. It could be argued...

---

⁴ For instance, if I ask you to “pass me the mug” and you answer “which cup?”, did you explicitly diverge and use a term you wouldn’t normally use (for instance, to express annoyance that I always ask you to pass me stuff), or did you not converge and use the term you would normally use (e.g., all cups are on saucer so you can see no mug)?
that, as social factors work as a modulatory factor of an underlying automatic mimicking tendency, they can only reduce but never lead to full divergence. The audience-design accounts are vague on the issue but we can hypothesise that as convergent linguistic behaviour is a signal of grounding towards mutual understanding, reduced convergence may be a sign of (potential) mis-communication or communication difficulties (and so may be individual cases of failed convergence).

The IAM claims that linguistic convergence is direct and automatic but, we can assume, not inflexible; inflexibility would mean that speakers’ linguistic choices are at the mercy of the priming mechanism and interlocutors would always inevitably converge on each other’s linguistic choices. We can identify two ways linguistic convergence can fail or be reduced under the IAM. First, on an individual occasion, Pickering and Garrod recognise that a speaker may consciously decide not to converge if she realises that convergence of linguistic behaviour does not correspond to higher-level alignment (e.g., semantic; she and her interlocutor are using the same word but for different referents). Thus, unlike convergence (which is direct and automatic), cases of failed convergence are strategic (and probably deliberate) and involve explicit audience-design considerations. They are only invoked as a separate repairing tool aimed at communication efficacy; as such they should require extra cognitive resources: speakers must in fact suppress the automatic mechanism to converge. This is in line with other two-stage accounts of language processing in interaction (e.g., the monitoring and adjustment model, Epley, Keysar, Boven, & Gilovich, 2004; Horton & Keysar, 1996; Keysar, Barr, Balin, & Brauner, 2000), whereby interlocutors process language on the basis of their own processing needs, and only if necessary, at a later stage, they engage in more sophisticated audience-design processing.

Note that, in contrast to the IAM, there is no reason to assume that failed and reduced convergence are computationally more costly than convergence for the social accounts. Even when they are assumed to be goal-driven (e.g., proactive route), both are functional tools to meet (opposite) social goals (Giles & Powesland, 1975) and social goals do not necessarily require deliberation (e.g., Lakin & Chartrand, 2003; see 2.1.2). Furthermore, the proactive social mechanism, in line with the audience-design accounts (yet in contrast to the IAM), seems to assume that high-order cognitive plans (like social goals and partner-specific
considerations) act at the very start of the production process (i.e., at utterance planning level, before referent selection and convergence/non-convergence on linguistic choices, cf. Giles et al., 1991; Clark, 1986)

Second, the IAM implicitly requires interlocutors to process each other’s speech in order for interpersonal priming to take place. Situations that prevent speakers from fully attending to the other’s speech (e.g., simultaneous engagement in another task) may reduce the priming effect resulting in reduced convergence and poor communication success. This is an important point. It could be argued, in fact, that social factors (cf. modulatory social route) may similarly integrate with the automatic priming mechanism: they modulate the strength of activation of the processed linguistic material (e.g., in the mental lexicon for word choices) and so the priming effect (i.e., the likelihood to re-use that material). As also suggested by the modulatory social route, positive social factors may increase speaker’s attention to their interlocutor’s incoming speech; the increased attention would translate into stronger activation of the linguistic representations and, consequently, bring to a higher likelihood to re-use the corresponding linguistic items (i.e., stronger convergence). Branigan et al. (2007), for instance, found stronger convergence effects at a syntactic level when the participants were the direct addressee of the previous (prime) utterance rather than a mere by-stander (and the utterance was addressed to someone else). Arguably, being the target of someone else’s speech made participants process that speech more deeply, leading to stronger convergence effects. Equally likely though, not being the target of the speech could have made by-stander process it more shallowly, leading to weaker priming effects. Negative social factors could work similarly to lack of attention: they decrease the strength of activation and the subsequent chance to re-use the linguistic material. However, while it is easier to see how the IAM can accommodate why and under what circumstances speakers may not converge (ad-hoc suppression of automatic tendency) or converge less (reduced activation of linguistic representation), it is less straightforward to see whether the IAM really leaves room for activation and priming to be enhanced (by positive special factors) given its claims that priming is already a pervasive mechanism. Is maybe “enhanced convergence” limited to how quickly interlocutors converge with each other or to the range

\footnote{If so, we would find that the magnitude of the priming (and so linguistic convergence) would act as a mediating variable between diminished attention and communicative success. This possibility, albeit interesting, has not yet been tested in the literature.}
of the convergent linguistic aspects? Furthermore, always under the IAM, communicative goals can induce speakers to strategically suppress convergence; can they also strategically enhance it? The strategic use of non-convergence happened at a later stage within the IAM, would the strategic enhancement of convergence also happened at a later stage (and be similarly associated with cognitive costs)? The questions and considerations raised in this section have no easy answer, and evidence is lacking as to how the automatic mechanisms assumed by some theories as the building block of linguistic convergence may interplay with social factors in modulating speakers’ tendency in conversation to lexically (and linguistically) converge. In Chapter 3 and 5 we address these issues and experimentally manipulate the social context of the speakers for then measuring how this affects their propensity to lexically converge with their interlocutor’s lexical choices.

Several questions remain open. They relate to the underlying issue as to what extent the proposed multiple mechanisms are compatible and how exactly they can interact. It has been proposed that linguistic convergence may be the result of a multicomponent mechanism (cf. Branigan et al, 2016), a suggestion that still needs to clarify how exactly these different components get orchestrated and work together. How do lower-level processes leading to linguistic behaviour matching (e.g., priming or implicit learning) unfold alongside more top-down processes like adapting to an interlocutor’s comprehension needs or a speaker’s own social goals? Is there a hierarchy between the mechanisms or a form of centralised “prior assessment” of the situation that prioritise one over the others (cf. Krauss & Pardo, 2004)? For instance, in the literature of common ground, it has been suggested that bottom-up mechanisms and higher-level information and plans may be simultaneously active. The extent to which each mechanism constrains speakers’ language processing (e.g., linguistic choices) is probabilistically determined by its moment-by-moment salience and validity respect to all the others (Hanna, Tanenhaus, & Trueswell, 2003; see also Barr & Keysar, 2006). Similar constraint-based probabilistic dynamics may govern linguistic convergence. The model, however, does not specify how exactly these probabilistic updates occur. In the next section, we discuss an account that may explain it.

As far as we are aware, there is no direct evidence for how the different mechanisms proposed for linguistic convergence may be intertwined. Nonetheless, Linnemann and Jucks
(2016) found that interviewees converged lexically more with interviewers who used a more schematic rather than a more elaborated language. Despite this, they liked the elaborated style more, showing that considerations of communication efficiency may overshadow affiliative ones. Using a triadic dialogue setting, Foltz, Gaspers, Thiele, Stenneken, and Cimiano (2015b) presented speakers with two sources of linguistic input: an addressee and instructions via headphones. Speaker and addressee named pictures for each other. Sometimes, the speaker heard different names for the same picture from their addressee and subsequently via headphones (e.g., “jug”/”pitcher”) before they named the picture themselves. Despite doing the task together with the addressee, speakers converged with the lexical choice of the most recent headphone input. Under these circumstance, where mutual understanding was not really at stake (i.e., addressees were able to identify the correct picture whatever label the speaker may have chosen), automatic mechanisms based on enhanced memory activation could overrule adaptation to the co-present addressee. And finally, Wang et al. (2015) found that online forum participants converged on each other's lexical choices when providing support but not when providing information. A possible explanation is that while providing support boils down to showing empathy and creating emotional bonds, providing information, if aimed at being effective, requires adding things that have not been already said; a task that requires new words. These latter results are important in that they show that communication is never at the mercy of linguistic convergence. Moreover, factors like interactional context and goals may significantly shape this pervasive tendency.

2.3 THE DYNAMICAL-SYSTEM ACCOUNTS

Dynamical-system accounts of dialogue (synergy model, Fusaroli et al., 2013) and, more generally, interpersonal interaction (coordinative structures model, Shockley, Richardson, & Dale, 2009), represent a recent theoretical and methodological development that draws on models in physics and biology (Kello & Van Orden, 2009; Kelso, 1995) and that has shifted the attention from the interacting individuals to the interaction itself (Riley, Richardson, Shockley, & Ramenzoni, 2011).

The approach acknowledges that a wide arrays of plausible cognitive processes have been suggested to underlie social and linguistic interaction (e.g., priming, Pickering & Garrod,
2004; explicit common-ground representations, Clark, 1996; memories for shared experiences, Horton & Gerrig, 2005a; perception-production coupling, Gallese, 2008; social goals and variables, Giles & Coupland, 1991). However, the complexity of humans' capacity to interact cannot be satisfactorily captured by only one or a subset of these mechanisms: dynamical systems would provide an integrative framework.

The account starts from the observation that linguistic interaction unfolds over time over many channels and, as such, should be best modelled as a time-evolving “complex system”. A complex system is the emerging product of interleaving mechanisms that, by influencing and mutually constraining each other in a direct and continuous manner, give rise to a coherent performance without the need of a centralised control (Richardson, Dale, & Marsh, 2014). The analogy often used is the one of a person striking a chisel with a hammer: the individual muscles and joints in her arm self-adjust and regulate each other in light of the external goal. This happens locally, without the intervention of a central controller that simultaneously represents and integrates current position and possible future directions of all the muscles/joints involved (self-organisation, cf. Bernstein, 1967).

This conceptualisation leads to two fundamental properties of interaction. First, every conversation is construed as realising a function: from explicit cooperative purposes like agreeing on what to eat for dinner, to more general goals like maintaining rapport by phoning a friend. This function acts as a guiding coordinating principle (just like the function of hitting the chisel guides the intertwining of the muscles and joints in the arm to best suit the ongoing goal); it is in light of this function and its demands that interactions must be investigated and understood (Fusaroli et al., 2013). Second, the many mechanisms that contribute to interpersonal interaction (e.g., priming, social goals) do not simply add up but operate in an interdependent manner so that each is constantly and continuously shaped by the others to optimise the underlying function (synergy, Dale, Fusaroli, Duran, & Richardson, 2013). A result of this process is that interacting individuals come to develop continuous mutual adaptation\(^6\). This adaptation does not only take the form of linguistic (and behavioural) convergence but includes compensatory complementary behaviours, and a flexible and constantly rearranging of labour division: interacting individuals become

\(^{6}\) Coordination in the dynamical-system terminology.
**dynamically coupled** (Hasson & Frith, 2016). Because of this coupling, the dyad is claimed to operate as a social unit; it is this unit and not its individual components (i.e., the conversation as a whole and not the individual contributions of the single interlocutors) that the study of social interaction must concentrate to gain an insight into the mechanisms underlying human communication (Fusaroli & Tylén, 2016; Marsh, Richardson, & Schmidt, 2009). This means that individual components (e.g., individuals and their linguistic/cognitive system), functional context (e.g., the goal of the interaction) and situational factors (e.g., social variables) are equal key players in the collective organisation of the “conversation system”; none of them in isolation can provide an appropriate understanding of it.

It is important to highlight here with regards to the discussion of linguistic convergence that the dynamical systems framework is not thought to supplant but rather to extend the theoretical accounts of convergence and coordination (Fusaroli et al., 2013; Fusaroli & Tylén, 2016). To reiterate, linguistic alignment and convergence are only one of the possible ways in which interacting individuals can become coupled (Hasson & Frith, 2016). In contrast to the IAM, however, they are not the centrepiece of our interactive ability: they are necessary but not sufficient. Complementarity behaviours (e.g., mechanisms like turn-taking, answering when asked a question) and procedural routines (e.g., sequential procedures of steps to solve a problem) become key elements to measure our interactive ability and its efficacy (Fusaroli et al., 2013; Mills & Gregoromichelaki, 2010). Alignment, restricted to task-relevant vocabularies (e.g., Fusaroli et al., 2012; Nenkova, Gravano, Hirschberg, & Gravano, 2008; Foltz et al., 2015a), is considered an "igniting tool" able to quickly establish task-relevant common ground and procedural routines but whose importance decreases over time (Fusaroli & Tylén, 2016; see also Mills, 2014 for a discussion of procedural coordination and complementarity).

### 2.4 Types of linguistic repetitions

Having reviewed the main mechanisms proposed for linguistic convergence, in this section we address two important distinctions: the distinction between communicative interpersonal convergence and non-interactional inter-speaker repetitions; and the distinction between short-term and long-term convergence. These distinctions will become important in Chapter 5, where the analysis of lexical convergence in unwieldy spontaneous dialogues (albeit lab-
based) made possible by recent methodological developments from dynamical systems will face the challenge of how to distinguish them.

### 2.4.1 Interpersonal convergence, topicality and linguistic conventions

The way we have so far discussed and conceived linguistic convergence is two speakers becoming linguistically similar to each other (e.g., using the same words) when talking to each other because they are mutually influenced by the other’s linguistic choices (as the automatic-cognitive and dynamical-system accounts would explain) or because they accommodate the other out of communicative (audience-design account) or social (proactive social account) goals. This is what we can further define as *communicative interpersonal linguistic convergence*: I use “*mug*” instead of “*cup*” (or anything else) because you used “*mug*”. There are situations that facilitate the repetition of linguistic material between two speakers, especially at lexical level.

For instance, you are buying ice-cream and ask for “*hazelnut, pistachio and vanilla*”. Time for the vendor to grab a cone and he has already forgotten what you wanted so he asks “*pistachio and?*”. You both referred to the greenish ice-cream flavour as “*pistachio*” because this is how conventionally an ice-cream made of pistachio nuts is referred to and there aren’t any conventional alternatives. Similarly, if you discuss which European capital to visit next with your partner, you may both frequently mention places like *Lisbon*, *Stockholm* and *Athens* over the course of the heated discussion. And if you’re guiding your child into the secrets of perfect lasagne making, you may find you’ll both be talking about *layer* and what comes *next*. Often, interlocutors’ syntactic and lexical choices are driven by linguistic conventions (“*pistachio*”), topic of the conversation (“*Athens*”) or the task at hand (“*next*”), all factors that restrict the variability in lexical choices and explain many cases of lexical repetitions between the speakers. For instance, interlocutors do not usually jump from topic to topic in closely adjacent turns (Grosz & Sidner, 1986) and conversations tend to display a high level of topic clusters as a result (Reitter & Moore, 2007). Note that audience-design accounts would explain these cases of convergent behaviour as separate from repetitions due to conceptual pact or partner-specific considerations: their occurrence is not brought about by previous use by the interlocutor but by the fact that the repeated word was the most probable choice (e.g., there is no other common way to refer to Athens; cf. Brennan & Clark,
1996). And while a by-product of the conversation (it is because you are discussing Athens that you repeat “Athens” several times), such cases of lexical repetitions cannot be explained in terms of interactional priming by the IAM but are rather top-down driven by meaning and situation model; nonetheless, they still constitute a sign of alignment of semantic representations and situation models 7.

2.4.2 Short-term vs. long-term linguistic convergence

Another important distinction is the one between short-term (or local) convergence and long-term (or global) convergence (also termed ‘adaptation’). Short-term convergence tracks those repetitions between interlocutors that take place in closely occurring utterances, like when you ask your partner to pass you the mug and they reply “which mug?”.

Long-term convergence represents adaptations that can be found far apart in the course of a conversation (even over minutes). In the literature, we encounter two conceptualisations of long-term convergence. Within the IAM, authors have talked of ad-hoc routinisations (Pickering & Garrod, 2006): recurrent combinations of words that get stored in and retrieved from memory as a single lexical representation. They usually indicate that interlocutors have converged on common higher-level linguistic representations (semantic, syntactic and phonological); for instance, they have developed a common reference system, like ad-hoc conventions established between them. Many partner-specific convergence results discussed above (e.g., speakers keep referring to “the heel shoe” with the same interlocutor even if the adjective is no longer necessary) can be explained in these terms. Once established, these routines are persistently re-used over the course of the conversation. Note that this form of long-term linguistic persistence should not be confused with increased short-term

7 There are more extreme cases of people’s language use becoming similar for non-interactional reasons. Imagine two couples admiring the Mona Lisa portrait at the Louvre (at different times). Both couples will probably talk about Leonardo, the size of the painting, the landscape in the portrait, and Tuscany, and the famous smile, and who the lady might have been, etc. Not just the two dialogues but also the linguistic production of the members from these different couples will show rather high levels of lexical similarity (even if independently, they all talked about the same thing which also happened to be a tangible element of the environment on which they are spending attention). However, their level of linguistic similarity is surely not due to coordination dynamics arising from communication with the other person. Note that according to the IAM, people who share highly activated semantic and mental model representation also for non-interactive reasons (e.g., they looked at the same portrait) are likely to use the same linguistic expressions in the same way.
convergence over time (i.e., interlocutors increasing the extent with which they repeat each other’s linguistic choices turn-by-turn as the conversation unfolds).

Others have suggested that short-term and long-term convergence may be underlined by qualitatively different processes (Ferreira & Bock, 2006; Reitter & Moore, 2007). While short-term convergence could be explained in terms of priming mechanisms (i.e., transient activation of linguistic representations), long-term convergence would be better accounted for in terms of implicit learning mechanisms (cf. Chang et al., 2006): repeated exposure to a certain structure that persistently increases its availability and use.

2.5 THE OPERATIONALISATION OF LINGUISTIC CONVERGENCE

Various measures of linguistic convergence (or alignment) have been devised in experimental and corpus studies to test and corroborate the different accounts of linguistic convergence in interaction (cf. Xu & Reitter, 2015). In this section we review the main operationalisations with the goal of understanding what properties they ascribe to linguistic convergence and the underlying cognitive mechanisms, in light of the different characterisations highlighted in the previous section. We limit our focus to linguistic convergence (lexical and syntactic) as opposed to convergence on non-linguistic aspects of speech (e.g., pitch), as it more closely relates to the experimental work in this thesis.

2.5.1 Laboratory-based operationalisations

Laboratory-based findings have mostly relied on task-constrained conversations, usually using a variation of the referential communication task. Here, pairs of participants communicate information to each other in order to complete a task that they couldn't otherwise complete alone (e.g., arranging an array of objects according to a certain order, identifying an abstract figure, guiding each other through a maze). Such task-oriented dialogues enable researchers not only to analyse what interlocutors say but also, by observing how they act upon each other’s instructions (e.g., what they look at or reach to grab), to obtain evidence of what speakers meant and what addressees understood (Brennan et al., 2010). Paradigms vary in the extent with which they constrain the conversation. Participants were more (e.g., Garrod & Anderson, 1987) or less (e.g., Branigan et al., 2000) free to contribute to the dialogue, they were given fixed (e.g., Brennan & Clark, 1996) or
interchangeable roles (e.g., as instruction giver and receiver; Haywood, Pickering, & Branigan, 2005), and interacted with another naive participant (e.g., Garrod & Doherty, 1996) or a confederate (e.g., Branigan et al., 2007). In this context, linguistic convergence has been measured through two manipulations.

**Choice for unconventional referents**

As we saw in section 2.1.3, one way to measure lexical and semantic convergence is having participants describe complex visual stimuli (e.g., abstract tangram figures, maze paths) that have no conventional name or denotation. Participants do this over several trials. The underlying logic is that the less strongly linguistic conventions determine how to refer to an object, the more room there should be for communicative interpersonal convergence (cf. Foltz et al., 2015a). Indeed, studies using this type of set-up found that participants begin with proposing complex interim denotations and, as the activity unfolds, converge on shorter and more effective expressions for the same stimuli. Thus, the manipulation successfully showed that if people must communicate about ambiguous referents that require a great deal of conceptualisation in situations in which they cannot rely on joint visual attention (e.g., participants usually do not see each other or each other’s stimuli so, for instance, they cannot point), appropriate labels for those referents are negotiated over the course of the conversation and eventually established. The manipulation is thus suitable to investigate long-term convergence as routinisations. Furthermore, this manipulation has so far only been applied to cooperative situations in which it was essential for participants to understand each other in order to accomplish a common task, and have mainly been used to investigate lexical/conceptual processes.

**Choice between two alternatives**

The second experimental manipulation consists in introducing (e.g., the names “chair” and “seat”) or syntactic (e.g., “the nurse gives the pirate a cake” vs. “the nurse gives a cake to the pirate”) alternatives, one of which is overall preferred or slightly more common. This is the logic behind all interactive priming paradigms: the experiment involves manipulating the occurrence of a specific linguistic characteristic (e.g., suitable-yet-dispreferred names for
objects) in the language of a naïve or confederate interlocutor (via scripted speech) and measuring the frequency with which the participant repeats this characteristic. For instance, in Branigan et al. (2011) participants took turns at naming and matching pictured objects. Some pictures had a highly favoured name (e.g., the name “chair” for the picture of a chair) but could also be correctly denoted with an alternative dispreferred name (e.g., the name “seat”). These pictures were always named first by the participant's (real or apparent) dialogue partner using the correct-yet-dispreferred name. After a number of trials it was the participant's turn to name the same picture; the researchers measured how many times participants converged on the dispreferred name introduced by their partner, under the assumption that they would only do so because their partner had done so (and that they would have otherwise used the preferred alternative). The empirical advantage of this manipulation is that it exploits the corroborated phenomenon that people tend to converge more on less frequent linguistic material (cf. “surprisal effect”; Jaeger & Snider, 2013). It thus manages to detect rather small effects in rather small samples (e.g., Branigan et al., 2011). Importantly, the fact that the lexical (or syntactic) choices introduced by the dialogue partner are unusual or slightly less common, and that the same picture is never named more than once by the participant allows the manipulation to limit the “inflation” of convergence rate driven by linguistic conventions or topic. For the same reason, it can be used to track short-term (turn-by-turn) convergence and whether this increases or decreases over time.

As a consequence of the manipulation, however, studies can only look at the linguistic behaviour of a single interlocutor and not of the dyad as a whole. In addition, even if alternated with fillers, this manipulation together with the repetitiveness of the task (usually, picture naming/matching) cannot rule out that lexical (or syntactic) decisions can become evident to participants. Thus, while spontaneous, convergence could be the result of participants’ conscious decisions about their lexical (or syntactic) choices (with the result that outcomes may be biased by strategic considerations).

A proposed alternative is to focus on the convergence of function words (e.g., pronouns, prepositions, conjunctions; Niederhoffer & Pennebaker, 2002). Function words are meant to capture the style of communication of a speaker but not the content of this communication; unlike content words (e.g., nouns), they would be independent of conversational topics. For
this reason, their production should be under less intentional control by the speaker (Tausczik & Pennebaker, 2010). Using a combination of hand coding and automated semantic-analysis (i.e., Linguistic Inquiry and Word Count (LIWC); Pennebaker, Francis, & Booth, 2001), the authors developed a measure of convergence (Language Style Matching (LSM); Gonzales et al., 2010; Jones, Cotterill, Dewdney, Muir, & Joinson, 2014) which tracks the covariation of the rate of use of various function-word categories in the interlocutors’ turn-by-turn production or across the whole conversation. Using this measure, Taylor and Thomas (2008) found that similarity in the distribution of articles, propositions and negations was a good predictor of the success of real-life negotiations (see also Ireland et al., 2011; Richardson, Taylor, Snook, Conchie, & Bennell, 2014). However, this is a very coarse operationalisation of lexical convergence that only captures the similarity in the distribution of word categories but not the temporal dimension of the convergence (paradoxically, speaker A: “I ate at 9 PM after sunset”, speaker B: “ate PM sunset after 9 I at” and speaker C “I slept by 3 o’clock before lunch” would score high on LSM but it is hard to see how it could fit the understanding of communicative interpersonal coordination).

2.5.2 Computational operationalisations

New statistical techniques have led to new operationalisations of convergence that do without choice between alternatives and priming manipulations to define lexical choices, and that can measure linguistic convergence in less linguistically constrained settings and outside purely referential contexts. Furthermore, many of these measures can track convergence in both interlocutors at the same time.

These operationalisations were generally introduced in computational linguistics and applied to corpus data of task-oriented dialogues (e.g., Reitter & Moore, 2014), natural conversations (Howes, et al., 2010), or a mixture of the two (e.g., Fusaroli et al., 2012). Some measures mimicked laboratory-based operationalisation and tracked only specific alternative structures (e.g., Howes et al., 2010), while others tracked all structures/ words (e.g., Reitter, et al., 2006). Although several new techniques have been proposed, in this thesis we were interested in exploring the potential of one of these techniques, Cross Recurrence Quantification Analysis (CRQA), for tracking lexical convergence in unconstrained lab-based conversations. Our choice was driven by the desire to measure convergence in
dialogues that, despite being experimentally manipulated, were free and unconstrained. The unconstrained dialogues produced by the participants in our Experiment 4 (Chapter 5), like spontaneous conversations, unfolded over time in unforeseeable yet not fully random ways; lexical production followed non-stationary and non-linear patterns. CRQA could not only address these issues but also allow us to quantify the magnitude together with stability, variability and predictability of lexical convergence over time. Before addressing CRQA in detail in Chapter 4 and 5, in this section we provide a review of the main new operationalisations proposed.

A first family of operationalisations is based on text similarity (as either string-based or semantic-based). Dale and Spivey (2006) used cross-recurrence analysis (CRQA) to quantify the degree of syntactic convergence (as word-class bi-grams) in dialogues between adults and children, while Fusaroli and Tylén (2016) used it on morpheme- and prosodic features. CRQA, as it is explained in detail in Chapter 4, operationalised convergence by treating interlocutors' speech as two discrete time series and estimating the amount of repetitions (or co-visitations) between the two series over time. Co-visititation is identified if interlocutor B repeats a word previously used by interlocutor A (and vice-versa). The advantage of CRQA is that it not only provides measures of the magnitude of convergence (i.e., the percentage of times speaker B's utterances share the same patterns of words as speaker's A utterances at a later point in the conversation, and vice-versa), but also its variability, stability and complexity. It allows tracking time-dependent and leading-following dynamics and information about the maximum lag of coordination. Finally, it can easily track convergence in open-ended spontaneous conversations. Note, however, that without ad-hoc adjustment, CRQA tracks both short-term and long-term convergence in one single measure.

Another similarity-based operationalisation uses vectors to represent (e.g., lexical aspects of) speech and distance in vector space to measure linguistic similarity (and so convergence). For instance, Foltz et al. (2015b) constructed for each speaking turn, a vector containing the frequency of all open-class words used (e.g., “Move the green circle left. Place the green triangle under the green circle.” would correspond to vector \([\text{move} = 1, \text{green} = 3, \text{circle} =\])

---

8 This literature is broad and we had to inevitably restrict the scope of the methods reviewed. In particular, we will not discuss the relevant but not critical complexity matching measure recently proposed by Abney, Paxton, Dale and Kell (2014) for reasons of limited space and time.
Convergence was measured as the cosine similarity of the word-frequency vectors between interlocutors’ adjacent turns; in simple words, the degree (but not the order) with which the same words were used to the same extent between interlocutors. As with LSM, the method coarsely captures the similarity in the distribution of words between speakers but lacks temporal resolution.

Importantly, both CRQA and cosine similarity track repetitions between speakers, whatever their origin (e.g., the language of two persons talking about the Mona Lisa at different times and with someone else will show high level of linguistic similarity because of the topicality effect and not because of interpersonal priming or partner-specific or social mechanisms). However, a careful choice of baseline against which to compare the empirical measure could help at least estimate the amount of linguistic convergence that occurred on top of topicality and task. A suitable option for task- and topic-constrained experimental dialogues is to compare the observed convergence against the convergence obtained for pseudo-pairs (i.e., pairs of people who did not actually interact; Foltz et al., 2015b). In this specific scenario, in fact, pseudo-pairs retain the repetitions that could have happened because of conventions or task-demand, but not because of interpersonal priming or other accommodation mechanisms (that is, because dyads were imposed the same task by the experimenter and were engaged in the same conversation topic, random conversations should conserve distributional as well as temporal properties as the original conversations, but not the coordination dynamics between interlocutors). Both methods can be extended to track time-dependent trends by the use of sliding windows (e.g., Liebman & Gergle, 2016).

Another type of operationalisation is based on re-occurrence probability. This operationalisation requires identifying a prime and a target between speakers' production (cf. Xu & Reitter, 2015) and deciding in advance the set of linguistic items (e.g., prepositional vs. direct-object structures) for which to track convergence. Church (2000) and Dubey, Keller, and Sturt (2005) developed a method to measure lexical and syntactic similarity in dialogues by dividing dialogue transcripts into halves: a preceding prime and a temporally-subsequent target. Convergence was operationalised as re-occurrence probability: the frequency of occurrence of a word (or syntactic rule) in the target given that it also occurred in the prime, normalised by the frequency with which the word appeared in the target
regardless of whether and how often it was in the prime. Prime and target do not have to be defined in terms of halves of dialogue transcripts (as in Church’s original study); for instance, the method was extended so that the prime was a speaker’s utterance and the target her interlocutor’s immediately subsequent utterance, and used to estimate short-term lexical convergence (i.e., the probability of a certain word occurring in interlocutor A being followed by the occurrence of the same word in interlocutor B’s subsequent utterance; e.g., Danescu-Niculescu-Mizil et al., 2011; see also Fusaroli et al., 2012; Garrod & Doherty, 1994 for similar turn-by-turn re-occurrence probability measures).

Reitter and Moore (2014) developed this method further to track (short-term) syntactic convergence as a temporal-decay effect: the probability that a certain syntactic rule occurs in speaker B’s utterance (target) as a function of the distance (in number of utterances or seconds) from a previous occurrence of the same rule in speaker A’s speech (prime). Linguistic convergence corresponds to a negative effect of distance/decay rate on repetition probability: the further away from the prime, the less likely we should be to find a repetition of the same item (see also Reitter, Moore, & Keller, 2006). The same approach has been effectively adapted to measure lexical and prosodic-features convergence (Ward & Litman, 2007b).

A final note: discussing how linguistic convergence is operationalised is not a pure exercise in style. A specific operationalisation can influence results. For instance, Howes et al. (2010) analysed the extent of convergence on the alternative dative constructions in a corpus of natural dialogues. They assigned a score of 1 to a sentence if and only if this contained the same dative alternative as the closest prime sentence, 0 otherwise. In a first measure, they calculated a mean score of similarity/matching for each dialogue partner (by summing up her matching scores and dividing it by the number of sentences). When compared against chance (i.e., the average similarity scores obtained for pseudo-interaction and shuffled-data), real interactions did not show higher average similarity. However, when the authors operationalised alignment as the likelihood to match (vs. not match) a previously encountered structure via logistic regression, and so considered every occurrence of an alternative construction as a data point (instead of the average-based approach just outlined), they found that interlocutors’ convergence tendency was above chance. Therefore, research
on linguistic convergence needs studies that compare how the different operationalisations perform and correlate on the same set of dialogues, and whether their results reflect predictions from theoretical models in psycholinguistics and cognitive science.

To conclude, laboratory-based measures of linguistic convergence have heavily constrained the nature of the linguistic exchanges; they have mostly focused on referential choices and imposed on the interlocutors either conceptually-complex referential puzzles or choices between alternatives. While valuable, these operationalisations strongly reflected the theoretical accounts they were meant to test. Analytical techniques developed in computational linguistics allow theoretically-free operationalisations of convergence; these operationalisations can measure linguistic convergence in less linguistically and conceptually constraint settings, and outside purely referential contexts, while tracking the linguistic behaviour of both interlocutors at the same time.

2.6 FUTURE DIRECTIONS

2.6.1 Beyond cooperative truthful interactions

The accounts of lexical (and more generally linguistic) convergence reviewed in the preceding sections seem to share the view that interpersonal convergence is a sign of good communication between the speakers (albeit for different reasons). However, studies that look at the automaticity of the priming effect (both lexically and syntactically), or that showed partner-specific audience-design modulations, or that more generally found that the magnitude of convergence positively correlated with interlocutors’ communicative success were biased toward very specific type of interaction: task-oriented, truthful, cooperative and either affect-neutral or positively valanced.

The experimental work had participants engaged in task-oriented dialogues (e.g., communicative referential tasks) where the communication between the participants was essential for completing the task at hand. In this type of scenario, speakers believed what they were saying and could expect the other to be doing what they seemed to be doing. Interlocutors were honest and "transparent", in that their goals were shared. This cooperation-bias in the study of interaction reflects the real-life expectation that in face-to-face communication people act cooperatively and transparently, and that they tell the truth
(Grice, 1975). It is findings under these specific conditions that most theoretical claims are based on. It is thus an open and important question whether we would still observe linguistic (in particular lexical) convergence in contexts other than the collaborative truthful interactions usually studied, and whether levels of linguistic convergence would be comparable.

In Chapter 3 we ask the question of whether and how a speaker's deceptive intents may integrate in and affect the processes underlying lexical convergence in dialogue. When people deceive, they have private interactional goals that are at odds with their interlocutor's expectation of cooperation and that clash with the goal which they ostensibly pretend to share. We leave to Chapter 3 the details of how a speaker's deceptive intents may affect lexical convergence. In the remaining part of this section, we take a more general stance and review evidence that more generally suggests how interactional contexts that do not encourage cooperation and/or rapport between interlocutors may affect linguistic convergence.

Let us start by admitting that different interactional contexts have been largely understudied, at least from a psycholinguistic perspective. The main accounts of linguistic convergence are also rather vague in this regard. Even when they superficially acknowledge that different types of interactional contexts, especially those involving an asymmetry in goals between interlocutors (e.g., competing and argumentative), may display different types of coordinative dynamics (e.g., Fusaroli et al., 2013), they do not explain why and how, nor do they make any clear predictions.  

Indirect evidence that communicative contexts other than truthful/cooperative ones may differently affect the magnitude of linguistic convergence comes from work on nonverbal convergence (or motor mimicry). However, the picture is far from clear. Paxton and Dale (2013) had pairs of participants freely discuss topics on which they agreed (e.g., movies they

---

9 For instance, we can assume that the IAM, for which linguistic convergence is essential for intelligibility and mutual understanding, would predict some degree of linguistic convergence also for non-cooperative non-truthful interactions. In fact, even conflicting interlocutors or speakers who are deceiving their partner may need a certain agreement on what it is at stake, an agreement that, according to IAM, is achieve via priming (more on this in Chapter 3). At the same time, however, two people disagreeing have less in common at their situation model, and this should relate to less semantic and lexical alignment, and so less convergent lexical behaviour.
both liked) and topics on which they disagreed and about which they had to convince each other. They found that argumentative dialogues disrupted nonverbal convergence compared to cooperative dialogues. In contrast, Tschacher, Rees, and Ramseyer (2014) found no detrimental effect of nonverbal convergence when participants interacted in a competitive (i.e., interlocutors had to argue against each other) rather than a cooperative context (i.e., interlocutors had together to convince a third party).

Some indirect insights also come from the findings within the social account framework that have linked disaffiliation to reduced or lack of convergence. Specifically, as reviewed above, when individuals find themselves in situations that they perceive as hostile or antagonistic, they tend to distance themselves verbally. This effect is sharpened when there is a clear outgroup/ingroup member dynamics (Bourhis, et al., 1991). At the same time, however, studies have shown that in case of asymmetries in the dominance of the interacting individuals, the individual in need of social approval tends to increase verbal convergence with their interactional partner (e.g., Putman & Street, 1984). And speakers were found to use high degree of linguistic convergence as a compensatory strategy to ensure smooth interaction in complex social situations (Balceitis & Dale, 2005; Louwerse, Dale, Bard, & Jeuniaux, 2012). The cloudy picture that results from these studies suggests that non-cooperative and non-truthful interactions deserve an experimental agenda.

Our personal stance is that investigating linguistic convergence outside cooperative truthful interactions is important, not just because evidence in this regard is lacking, but also because contexts other than cooperation (e.g. deception), apart from being rather common, often present a complex intertwining of cognitive and socio-affective processes. They can thus further our understanding of the interleaving between the different mechanisms proposed for linguistic convergence.

2.6.2 Social accounts: any evidence from lexical convergence?

There is another aspect of lexical convergence that we are concerned with in this thesis. As we saw in section 2.1.2, evidence in favour of social accounts mostly concerns non-linguistic aspects of language (para-linguistic levels). This evidence is mostly correlational and has difficulty clarifying the directionality of the effect. When experimental studies have tried to
clarify this directionality (i.e., whether positive/negative perception led to more/less convergence), they focused on interlocutors’ syntactic choices, used non-interactive settings, and obtained unclear results (Lev-Ari, 2015; Weatherholtz et al., 2014). Evidence at a lexical level has mostly concerned a coarse operationalisation of convergence based on the rate of use of linguistic non-content categories; a measure that, by itself, cannot capture more subtle and time-sensitive matching of lexical choices (and presumably corresponding semantic representations).

In Chapter 5, we investigate how the social factors that have been found to modulate convergence at para-linguistic levels (e.g., the use of convergence as a tool to manage social distance; the consequences on convergence of contextual social variables such as the impressions that speakers form of their interlocutor) may also affect interlocutors' lexical choices. We do this using a linguistically unconstrained interactive paradigm and exploiting the analytical characteristics of CRQA.

What could be expected at a lexical level? Can the findings on accent, pitch, and speaking rate generalise to what may happen at the level of word choice? And will lexical choices, unlike syntactic choices, provide a clearer pattern of results in line with the non-content linguistic literature? It is still unclear whether convergence on non-content speech attributes and convergence on content aspects of speech belong empirically and conceptually to the same phenomena. It is thus an open question whether social factors that were found to affect interlocutor’s tendency to converge on their partner’s non-linguistic aspects of speech may also affect their tendency to converge on their lexical choices.

Communicating is a multimodal process, with interlocutors simultaneously managing language at several levels (e.g., lexical, syntactic, prosodic, etc.), gestures, gaze and other behaviours like facial expressions. The IAM predicts that convergence at one level would reinforce converge at another level (at least in the linguistic domain), and there is indeed evidence that interlocutors can simultaneously converge on multiple channels of communication (e.g., Louwerse, et al., 2012). Nonetheless, dynamical-system and social accounts have suggested that linguistic convergence may not follow a unimodal pattern: speakers may be triggered by certain factors to converge at some levels (e.g., pitch) but not others (e.g., syntax) (e.g., Bilous & Krauss, 1988; Nenkova et al., 2008). As recently
suggested, it may be that “under certain communicative contexts not all aspects of communication are aligned equally and that the temporal convergence of some communicative structures carries more weight than others” (Paxton, Roche, Ibarra, & Tanenhaus, 2014, p. 5).

There are some obvious differences between a speaker's lexical choices and her accent, speech, or speaking rate. Some of these differences suggest that the social factors found to affect speaker’s convergence on the non-content aspects of their partner’s speech may not similarly modulate their tendency to convergence on their lexical choices. First, unlike other aspects of one's speech production (e.g., pitch), lexical choice is a linguistic domain where speakers enjoy many degrees of freedom and that rarely represents stable individual preferences (see Lev-Ari, 2015, for a similar observation on syntax). Words may thus be perceived as less discriminative of the individual speaker. As such, they may be less susceptible to the influence of social perception regarding convergence than other levels of language (e.g., accent or pitch) which, in contrast, represent relatively stable aspects of a speaker's linguistic production (Pardo, 2013). At the same time, lexical selection cannot simply be reduced to a mere stylistic choice; the words we choose can heavily affect meaning. So speakers may be less prone to accommodate out of social factors if this affects undermines the meaning of what they really want to say.

On the other hand, however, speakers tend to be more aware of the precise words they use compared to other aspects of their speech (e.g., syntax) and assume their interlocutors are too. This could make convergence on lexical choices sensitive to social-factor modulation as much as or even more than non-content aspect of speech. Lexical changes are in fact likely to be noticed by the conversational partner and so may be used by speakers for their social goals (as we saw, the proactive social route predicts that speakers adapt their linguistic style on the assumption that this can be perceived, cf. Giles et al., 1991).
Chapter 3

**LEXICAL CONVERGENCE IN DECEPTIVE INTERACTION**

### 3.0 CHAPTER OVERVIEW

In this chapter, we investigate how the processes that underpin real-time deception integrate in and affect the processes underlying lexical convergence in dialogue. When people deceive, they have private interactional goals that are at odds with their interlocutor's expectation of cooperation and that clash with the goal which they ostensibly pretend to share with the other. In three experiments and using two novel interactive priming paradigms we measured the extent to which deceiving speakers matched their interlocutor’s choice of words, and whether they did so to a greater or weaker extent than truthful speakers. We did not find any consistent evidence that deceiving differently affected lexical convergence rate from telling the truth.

### 3.1 INTRODUCTION

There is strong behavioural and physiological evidence that deceiving is a complex behaviour underpinned by a net of diverse cognitive sub-processes (e.g., Ganis & Keenan, 2009; Vrij, Fisher, & Blank, 2015). This complexity makes it cognitively more taxing than truth telling (e.g., Walczyk et al., 2005), which is in turn hypothesized to be a baseline mode of human behaviour (Gilbert, 1991; Levine, 2014; Verschuere & Shalvi, 2014). Violating this truth-bias makes deceiving also emotionally demanding (e.g., Ekman, 1985). Together, cognitive and emotional demands hinder the cognitive flexibility of deceivers (e.g., Leins, Fisher, & Vrij, 2012). However, deception and lying are not only taxing mental operations but also complex social activities that unfold in interaction (e.g., Sip, Roepstorff, McGregor, & Frith, 2008), and deceivers prove indeed to be skilful communicators (Buller & Burgoon, 1996).
The psycholinguistics of dialogue has so far limited its investigation to cooperative and/or truthful interactions (e.g., Garrod & Anderson, 1987; Fussell & Krauss, 1989). Nearly all we know about dialogue dynamics comes from studies that assumed that people believe in what they say and expect the others are doing what they seem to be doing; participants were honest, "transparent", in that their goals were shared. This cooperation-bias in the study of interaction reflects the real-life expectation that in face-to-face communication people act cooperatively and transparently, and that they tell the truth (Grice, 1975).

Findings from this line of research have shown that cooperative interlocutors tend to match each other's linguistic style (e.g., Brennan, et al., 2010), and that this interpersonal linguistic convergence is functional to effective communication (Pickering & Garrod, 2004; Reitter, et al., 2006) and a natural outgrowth of joint goals (Clark, 1996; Richardson, Dale, & Tomlinson, 2009). Interpersonal convergence has also been linked to the management of social outcomes, like fostering liking and trust (Giles & Powesland, 1975; van Baaren, Janssen, Chartrand, & Dijksterhuis, 2009). However, what happens when the circumstances change, and participants are no longer honest?

When people deceive, they have private interactional goals that are at odds with their interlocutor's expectation of cooperation and that clash with the goal which they ostensibly pretend to share. The question we ask in this study is how the processes that underpin real-time deception integrate in and affect the processes underlying linguistic coordination in dialogue. In other words, given that one partner's linguistic behaviour shapes another's during collaborative dialogue, how is this trend affected if one of the partner is deceiving the other? Do we still observe convergence of linguistic behaviour under conditions of no convergence of intents?

The question is interesting not just because we lack an investigation of deception from a psycholinguistics-of-dialogue perspective, but also because investigating deception, and its complex intertwining of cognitive and socio-affective processes, can further our understanding of the interleaving between the cognitive and social mechanisms of linguistic convergence.
Our focus is on lexical convergence. We start by drawing a definition of deception and lying. We then outline those characteristics of deception that may affect the lexical convergence tendency between deceiver and deceived. After revising key aspects of linguistic convergence theories and findings, we discuss how the tendency to lexical converge may be different in deceptive compared to truthful interaction. Finally, we present the experimental work.

3.2 DECEPTION AND LYING: AN ATTEMPT AT DEFINITION

Epistemologically, the question of what constitutes deception and lying is far from settled. According to one widely accepted characterisation, I lie to you if, knowing that ‘p’ is false, I tell you that ‘p’ is true in order to get you to believe ‘p’ (Jackson, 1991; Kupfer, 1982). Conversely, I intentionally deceive you if, knowing ‘p’ is false, I deliberately act in a way that induces you to believe ‘p’ (but without necessarily asserting that ‘p’ is true) (Jackson, 1991).

To clarify the difference, let's imagine we are work colleagues. I secretly despise you as you will probably be assigned the project in Provence that I hoped to get. Our manager just informed me that she is back in town and she wants to have a team meeting today at 2pm. I should let everyone know. I presume it will be an important meeting that will affect the decision concerning the project. When you ask me about the meeting, I tell you that you were inaccurately informed and that “there is no team meeting at 2pm”. I am lying to you and deceiving you into believing that there is no team meeting at 2pm whereas there is one. Now, let's suppose that there is no team meeting at 2pm, our manager just re-scheduled it for 11am. If I tell you that “there is no longer a team meeting at 2pm”, I am not lying: there is actually no meeting at 2pm. Yet my statement makes you hold a belief about something which is deceptive because you will wrongly infer that there is no team meeting at all today. Thus, a lie is always deceptive but not all forms of deception are lies, even if they share the same result: the deceived believes something untrue (Bakhurst, 1992).

The philosophy literature is full of attempts to refine these concepts (Arico & Fallis, 2013; Fallis, 2010), with people questioning the deceiver's real intents (Faulkner, 2007) or the epistemic implication of what's asserted (Carson, 2006). Unfortunately, we lack the space for
lingering over these speculations. Whilst acknowledging that there are some aspects that are controversial, we focus on the following property that all theories seem to agree on: deceiving and lying involve instilling (what you know is) a false belief in another person. In this study, we thus follow the characterisations above and operationalise deception as the deliberate attempt of inducing another person to believe something that you know is untrue. Thus, we construe deception as a speaker's interactional goal. We take lying as one specific realisation of deception: the speech act of deliberately stating something false as if it were true. In this chapter, we will use “deceiver” as general term and “liar” as the more specific term (for a deceiver who is telling a lie) when strictly relevant.

We acknowledge that these characterisations do not capture all the nuances of the phenomena under consideration. For instance, the characterisations do not distinguish the different intents with which people lie or deceive. We do not always lie to harm or take advantage of others. If you were invited for dinner by a new friend and, gosh, he really has issues with salt, the risotto is utterly insipid, when he asked you “How is it?”, you would probably answer “It's good!” White lies help us make our way through complex social dynamics and ease our communications; we won’t address them in our study. Furthermore, according to our characterisation above, lies entail stating something that does not correspond to reality, like “there is no team meeting at 2pm” when there is actually one. These are often defined as lies of commission and have received most attention in the cognitive sciences and are also receiving our attention in this study. However, sometimes we tell literal half-truths or truths that are taken out of context so that what is said or what is not said is misleading, like “there is no longer a team meeting at 2pm” when it is actually at 11am. Deceiving truths have received less experimental attention and are not addressed in our study. Finally, note that in openly-deceptive contexts, in which the others expect you might be lying (e.g., poker), a whole truth can be sold as a lie and fulfil the same deceptive intent (e.g., given you successfully bluffed before by betting heavily on a weak hand, you try to get your opponent into thinking you are bluffing again even if, this time, you are betting heavily on a truthfully strong hand).
3.3 CHARACTERISTICS OF DECEPTION AND LYING THAT MAY INTERPLAY WITH
LINGUISTIC CONVERGENCE MECHANISMS

3.3.1 A taxing affair

In this section, we discuss the findings that have contributed to the view of deception and
lying as multi-process mental operations that are cognitively more taxing than truth telling.
We outline the processes that are common to both lying and deceiving (and that distinguish
them from truth), and those that instead are peculiar to lying alone (but not of deceiving in
general). Although we do not yet discuss linguistic convergence in this section, the
properties of deceiving and lying outlined here will become important in section 3.5 when
we consider how linguistic convergence tendencies may differ in deceivers compared to
truth-tellers.

When people lie, they state something they know is not true. Behavioural and neuroscientific
findings suggest that this builds on three separate processes: a decision to lie, the suppression
of a dominant tendency to state the truth, and the construction of an alternative (false)
response. These are separate components in that they were shown to add to liars' cognitive
load independently, as measured in terms of increased response times (e.g., Duran & Dale,
2012; Walczyk, Roper, Seemann, & Humphrey, 2003; yet see Reynolds & Rendle-Short,
2011 for a critique), less stable and more complex arm movement trajectories (Duran, Dale,
& McNamara, 2010), and activation of brain areas associated with executive control and
working memory (i.e., prefrontal and anterior cingulate cortex, see Abe, 2011 for a review).

The decision to lie (or decision tout court?)

According to most cognitive approaches to lying and deception (e.g., the Activation-
Decision-Construction Model, Walczyk et al., 2005; Walczyk, Mahoney, Doverspike, &
Griffith-Ross, 2009), the default of our cognitive system is to produce the correct or
congruous response to a given stimulus. For instance, when people hear a question, the
relevant information is automatically activated and retrieved from episodic/semantic memory
into working memory. In light of their social/communicative goals, speakers assess the risks
and benefits associated with revealing that information. And decide whether to lie.
Recent studies have started exploring the decision-making costs of deceiving by creating situations in which participants could freely decide, within a deceptive context, when to lie or tell the truth (e.g., Duran & Dale, 2012; Eapen, Baron, Street, & Richardson, 2010), and using paradigms in which the participants' decision brought them gains or losses (e.g., Bhatt, Lohrenz, Camerer, & Montague, 2010). In these scenarios, studies found no difference in signs of cognitive cost (e.g., response times) when participants freely decided to lie compared to when they freely decided to tell the truth, suggesting that the demand of the decision process was not a by-product of lying itself but of the deceptive strategy more broadly (Carrión, Keenan, & Sebanz, 2010).

Other studies explored the difference between lying and deceiving by the use of double manipulations that separated the truth value of a statement (true/false) from the intention of the speech act (honest/deceitful). For instance, in Abe, Suzuki, Mori, Itoh, and Fujii's (2007) PET study, participants replied to four autobiographical questions (e.g., “What is the name of your primary school?”) and lied about two. The study used two experimenters: a first one instructed the participants about which questions to lie to; a second one told them to deceive the first one by doing the opposite of what he had asked them to (i.e., respond truthfully to the questions he had asked them to lie about, and vice-versa). The study found effects of deception both when participants dishonestly (i.e., unbeknownst to the first experimenter) told the truth and when they lied: longer response times, higher level of anxiety (self-assessed) and activation of the ventromedial prefrontal cortex and the amygdala (i.e., regions usually associated with emotional conflict resolution).

Together, these findings suggest that there is a cognitive cost that results not simply from lying but from all those situations in which people consider the possibility to deceive and assess the pros and cons of telling the truth. Note that these are quite common in real life, where the boundary between honesty and deception is often just a matter of seconds. The candid phone call with your long-distance partner can suddenly turn into a deceitful conversation when he starts inquiring about your last night out. Moreover, these findings also indicate that the cognitive effort associated with the decision-making process will show behavioural and physiological markers even when you eventually decide to tell the truth.
The suppression of a truth-bias

But you decided to lie. There is now all this prominent true information active in your working memory that needs to be concealed, a process that requires additional cognitive resources. For instance, using “yes/no question” paradigms and instructing participants about when to lie (tactics that single out truth-suppression from lie construction and the decision to lie), studies have found extra processing-load effects in the range of 200-300ms (Ganis, Kosslyn, Stose, Thompson, & Yurgelun-Todd, 2003; Spence et al., 2001; Walczyk et al., 2005) and larger medial frontal negativity in brain areas involved in monitoring and response-conflict resolution (Johnson, Barnhardt, & Zhu, 2004).

That inhibiting factual information is hard and does not always work finds support in the ironic-processes theory (Wegner, 2004), whereby a person is more likely to behave counter-intentionally when highly motivated to behave in a certain way. For instance, participants were more likely to refer to a piece of visual information when explicitly asked to conceal it (Wardlow Lane, Groisman, & Ferreira, 2006). Not surprisingly, this evidence suggests that the inhibition-effect, despite being present when we lie, is not necessarily unique to the deceptive act but may come together with any strategic activity. Most probably, it is the by-product of any conflicting-response resolution, regardless of the source of the conflicting information (e.g., lying vs. dual-task resolution, Seymour & Schumacher, 2009).

The construction of a lie

You managed to withhold the truth and must now fabricate a plausible lie (rarely in life do we get away with just a yes or no). Unlike past experience, which can be retrieved from episodic memory, and unlike generic factual information, which can be recalled from semantic memory or inferred from what we are perceiving, the content of a lie has less or no scaffolding in the external world, and no entry in our long-term memory (at least if it is the first and spontaneous telling of the lie10): a lie is made up. Some studies have tried to investigate the cognitive cost individually imputable of lie construction. For instance, liars

---

10 Research has shown that tailored rehearsing (Vrij et al., 2009), strategic planning (Vendemia, Buzan, & Green, 2005) and simple practice (Hu, Chen, & Fu, 2012) make lying easier and minimize (e.g., Johnson, Barnhardt, & Zhu, 2005; Verschuere, Spruyt, Meijer, & Otgaar, 2011) or cancel out (Vrij, Granhag, Mann, & Leal, 2011) behavioral and physiological difference between lies and truths.
took roughly 400ms longer to falsely reply to open-ended than yes/no questions, a delay that was ascribed to the fabrication of the false response (Walczyk et al., 2005).

Other results corroborate the view of lying as a constructive process that depends on people's ability to access and manipulate linguistic information. When answering open-ended questions about personal data, people with higher verbal fluency skills were found to be faster at telling lies than people with lower skills, but not when they told the truth (Walczyk et al., 2003). Verbal fluency is a measure of effortful semantic memory retrieval that reveals people's ability to self-initiate, self-monitor and inhibit responses (Henry, Crawford, & Phillips, 2004). Nonetheless, note that different types of lies may be more or less demanding to fabricate. Factors like recency of the events (Abe et al., 2006), complexity of the task (Williams, Bott, Patrick, & Lewis, 2013) and emotional detachment (Walczyk et al., 2005) facilitate or hinder their production.

**Cognitive (in)flexibility**

Granted that deceiving involves additional cognitive processes than truth-telling (at least when truth-telling is not the result of the decision not to lie), and that the output of these processes (e.g., RTs) may differentiate between the two, recent studies compared the cognitive flexibility of truth-tellers and liars. The lack of direct experience of the lied-about fact made liars' stories less consistent and liars less able to manipulate the details of their stories (Leins, Fisher, Vrij, Leal, & Mann, 2011). Leins et al. (2012) asked participants to lie about having been in a certain place and having engaged in a certain behaviour, and to report about this twice: either across different modalities (e.g., in words and by drawing) or in the same modality (e.g., words-words). Overall, liars were less consistent than truth-tellers across reports. More interestingly, liars were less consistent when they reported the fact in two different modalities than in the same modality, whereas truth-tellers were not affected by modality. What caused liars' difficulties was that, unlike truth-tellers, they lacked any rich memory representation of the “experience” and could not recall and manipulate its diverse perceptual aspects (see also Vrij et al., 2008). It must be noted, however, that this may be characteristic of lies not because they are deceptive but because they are fabricated. Studies on memories for experienced vs. imagined events found that imagined events contained
worse temporal and spatial information as well as fewer perceptual details (Johnson, Foley, Suengas, & Raye, 1988).

3.3.2 Other forms of load

In the previous section, we presented evidence that generating deceptive responses depletes the attentional and executive resources of speakers, as observed in increased response times and greater activations of executive function areas in the brain. This evidence comes mostly from studies that treated deception as the production of isolated speech acts whose content bore no communicative consequences for the speaker, and that often used experimental paradigms whose key manipulation was to induce a visual or memory-based conflict in the mind of participants.

As said at the very beginning, however, deceiving (including the special case of lying) are forms of communication, and as such, they take place in a specific context, for specific purposes, towards a specific person who is assumed not to be aware of our intent. What makes deceiving harder is not a simple matter of working memory load, conflict monitoring and response inhibition (cf. Sip et al., 2008). Deceptive interactions involve other mechanisms that increase the complexity of the behaviour and that are linked to its social and interactive dimensions.

The emotional stress of deceiving

When interacting with someone, interlocutors usually assume each other’s good faith: that they are actually doing what they seem to be doing, and that they believe in what they're saying. Acting against this cooperative social pact not only requires extra cognitive effort, as we have seen in previous sections, but it is also emotionally burdensome. Emotional implications such as fear of being caught, desire to be believed, feeling of guilt and complacency can add to the great deal of what deceivers must conceal (Caso, Gnisci, Vrij, & Mann, 2005; Paul Ekman, 1985). And findings from neuro-imaging studies confirm the involvement of brain areas associated with emotional conflict resolution (Baumgartner, Fischbacher, Feierabend, Lutz, & Fehr, 2009).
There seems to be great variability across individuals and situations in this respect. For instance, dominant extraverted people appeared to be better at hiding nervous behaviour (e.g., leg movements) when deceiving (Riggio, Tucker, & Throckmorton, 1987; see also Vrij, 1993), and individuals are more likely to lie to people they do not feel close to (DePaulo et al., 2003). Besides these differences, evidence suggests that fear of being caught is proportional to the stake, and that the sense of guilt increases proportionally to the feeling of empathy with the ‘victim’ and the recognition of how severely they will be affected by our lies (Ekman, 1997). Also, the likelihood of being caught and the consequent repercussion on the deceiver's social image may add to the emotional demand and the related psychological stress (e.g., anxiety and the need to hide it). For instance, liars showed weaker confidence in their own credibility than truth-tellers (DePaulo et al., 2003; Vrij et al., 2008). In a mock crime-interrogation study, “guilty” participants invoked the right to silence more often than “innocent” participants, as this would avoid them verbal confrontation (Kassin & Norwick, 2004). Consistently, speakers’ linguistic accuracy and responsiveness were modulated by how confident they felt about what they were saying (Smith & Clark, 1993).

### 3.3.3 Deception and lying as a form of communication

But the social dimension of deception is not only detrimental. Recent findings suggest that, as in any complex social interaction, deceiving requires a broad set of affective-cognitive processes that revolve around the ability to weigh the different possible courses of action and identify how these possibilities could be coordinated with the interlocutor (Sip et al., 2012). We are concerned with mechanisms like keeping track of others’ beliefs, monitoring their feedback to anticipate their verbal responses, building trust through pseudo-cooperation, and estimating the long-term consequences of the deception (Sip et al., 2008).

**Self- and other-monitoring**

The intent to deceive requires the deceiver to take an intentional stance on their addressee who is deliberately manipulated to hold a false belief about a certain state of affairs, and about the real intention of the deceiver (Lissek et al., 2008). Deceiving is thus likely to involve mentalising, an evolved psychological mechanism that is necessary to represent intentions and expectations in social interaction (Leslie, 1987).
And indeed, at one level, deceivers have been shown to keep track of what the addressee has been led to believe, in order to maintain internal consistency and external plausibility of their false story (e.g., Duran & Dale, 2012). We can hypothesise that the deceiver is forced to maintain two mental representations: the real one, and the false one that they shared as ostensible truth with the addressee. For instance, imagine a witness intentionally providing false details to an investigator. There are facts known to both of them (e.g., a man was shot dead in the Meadows in the morning). There are other facts that are only known to the witness and that she is trying to hide from the investigator (e.g., a woman killed the man; she was waiting for him behind a fence; that woman is her mum). The witness does not know what exactly the investigator knows (e.g., has anyone else seen the murder?). To meet her conversational goals (i.e., deceive the investigator and prevent suspicions falling on her mum), the witness must take into account and keep separate what she knows the investigator knows (i.e., where and how the man was killed), what she only knows (i.e., her mum did it), and the facts that she is fabricating and sharing with the investigator during the interview (e.g., she saw a guy with a long coat running away; he looked agitated). To succeed, the deceiving witness must keep separate the situation model of what truly happened, and the situation model of the fabricated facts she is sharing with the investigator.

The second monitoring process is oriented to the addressee's beliefs about the intention of the deceiver. It is an essential process to help deceivers get away with their action and meet their long-term goals (e.g., Sip et al., 2012) by adjusting their deceptive message when necessary (e.g., Carrión et al., 2010; Schweitzer, Brodt, & Croson, 2002). For instance, Bhatt et al. (2010) used a bargain task paradigm, in which buyers had the possibility to bluff on the real value of items. Buyers, who received a certain value for the items, suggested a price to the seller. Using the suggestion, the seller proposed a final price to the buyer. The game rules made it advantageous for buyers to propose smaller prices (especially when the value was high), and for sellers to accept higher prices. Strategic buyers tailored their behaviour by taking into account how the seller would interpret it (i.e., as an honest or dishonest price suggestion). Strategic deceptive buyers were able to guarantee for themselves a socially positive image (as honest buyer) through a continuous monitoring and updating of the seller’s beliefs and by inhibiting the impulse to communicate the easiest information.
The language of deception and the use of feedback

The majority of studies that have investigated the linguistic dimension of deception were interested in identifying specific markers of lying. Findings came mostly from the application of machine learning techniques to the analyses of chat-like (Zhou, Burgoon, Twitchell, Qin, & Nunamaker, 2004, for an overview) and speech corpora (e.g., Benus, Enos, Hirschberg, & Shriberg, 2006; Graciarena et al., 2006; Hirschberg et al., 2005). For instance, a text analysis of productions about a variety of topics and modalities (i.e., spoken, computer/written; Newman, Pennebaker, Berry, & Richards, 2003) reported that deceptive speech is characterised by lower degrees of linguistic complexity: low rate of words expressing exclusion (e.g., without), distinction (e.g., but) and evaluation, and a high rate of concrete verbs like motion verbs. This would reflect the fact that creating a false story requires cognitive resources that deceivers might not have. Deceivers’ stories are thus less complex and focus more on simple concrete events, and avoid details hard to remember like distinctions or evaluations. Deceptive speech was also characterized by fewer references to the self, indicating a tendency in deceivers to distance themselves from what said (Zuckerman, De Paulo, & Rosenthal, 1981).

In her theoretical framework, the Interpersonal Deception Theory (IDT), Burgoon was one of the first to encourage moving beyond the study of self-standing speech act and micro-behaviour and focus on deceivers' overall performance (Buller & Burgoon, 1996). Even if Burgoon’s interest was in lie detection, the broadly relevant idea put forward by her IDT is that deception is at heart a communication activity and, as such, highly interdependent, dynamic and adaptive. The dynamics and characteristics of real-time interaction should be central to any research of deception. Within this framework, findings suggest that deceptive behaviour is facilitated by higher degrees of interactivity. Burgoon, Buller, Floyd, and Viprakasit (1998) observed that liars engaging in highly interactive dialogue with their victim (i.e., instead of monologue-like attempts) were more successful at deceiving. Interactivity allowed liars to tune in with their partner and adapt their deceptive message accordingly in terms of content, fluency and nonverbal demeanour. Similarly, deceivers were found to be responsive to the communicative behaviour of their addressee and interpreted it as feedback on their own performance, with positive demeanour (high involvement)
interpreted as acceptance and success, and negative demeanour interpreted as suspicion and input for change (Burgoon, Buller, White, Afifi, & Buslig, 1999). In a computer-mediated-communication experiment (Hancock, Curry, Goorha, & Woodworth, 2007), liars used fewer causal explanations and negations (which could reveal inconsistencies in their narrative or trigger requests for clarifications) than truth-tellers, a tendency that was enhanced the stronger their desire to be believed. Their language also showed a higher degree of similarity with their interlocutor’s.

Other studies have explored the interactive nature of deception by applying Conversation Analysis to naturalistic conversations (e.g., reality shows) (e.g., Reynolds & Randle-Short, 2011). Importantly, results suggest that normal mechanisms of dialogue must be accounted for when investigating deceptive interactions (see also Bavelas, Black, Chovil, & Mullett, 1990). For instance, longer response latencies previously reported as cues to deceptive (vs. truthful) answers in laboratory studies (e.g., Walczyk et al., 2005; Sporer & Schwandt, 2006), in real deceptive interactions they were found to be indicative of denials or blame shifting independent of lies (Reynolds & Rendle-Short, 2011).

Methodologically, the evidence reviewed in this section confirms the relevance of investigating deception in dialogic contexts and focusing on the adaptation dynamics between deceiver and deceived. When doing this, we face the challenge to distinguish forms of linguistic adaptation that happen in any communication, and adaptations specific to the deceptive intent (e.g., to gain the trust of the addressee or the result of additional cognitive difficulties).

3.4 LINKING DECEPTION AND LINGUISTIC CONVERGENCE

3.4.1 Automatic-cognitive and socio-strategic accounts of linguistic convergence

In order to see how deception may affect lexical convergence, we must first review the mechanisms that have been suggested to underlie linguistic convergence. In this study, we focus on two of these mechanisms (and hereby only discuss the relevant aspects of these mechanisms; for more details, see Chapter 2). According to one major cognitive account proposed in psycholinguistics (the Interactive Alignment Model (IAM), Pickering & Garrod, 2004), linguistic convergence is the result of automatic mechanisms grounded in the
architecture of our cognitive system: the automatic re-use of transiently activated memory representations. In contrast, social psychology and sociolinguistics have linked linguistic convergence (and non-convergence) to socio-affective dynamics, and have interpreted convergence as a tool used to fulfil social goals and manage social distance (Giles & Coupland, 1991) or as a natural response whose magnitude can be modulated by positive (vs. negative) affective experiences or goals (cf. Dijksterhuis & Bargh, 2001).

According to the IAM, linguistic convergence is explained in terms of shared representations for our speech production and comprehension systems and the fact that memory retrieval is biased towards previously activated material. In dialogue, both production and comprehension systems are active at once, so that automatic priming mechanisms (i.e., tendency to pick up and reuse expressions produced or heard shortly beforehand) at lower processing levels (e.g., word choice) enhanced convergence at higher levels (e.g., semantic). This has three interconnected effects: we observe linguistic convergence between interlocutors; behavioural adaptations lead to alignment of linguistic representations which, in turn, induce alignment of situation models. Importantly, according to IAM, processes that reduce (vs. reinforce) speakers’ attention to and memory encoding of the incoming speech signal might hinder (vs. enhance) interpersonal priming and linguistic convergence (and thus alignment at higher levels). Ours is a speculation as there is no empirical evidence for this (but see Branigan et al., 2007; Weatherholtz et al., 2014).

Another influential account argues that speakers may adopt the linguistic style of their interlocutors (e.g., re-use their choice of words) as a way to manage social relationships and fulfil affective goals (e.g., Communicative Accommodation Theory, CAT; Giles, et al., 1987; Giles & Smith, 1979). For instance, they may use linguistic convergence to enhance affiliation, and induce trust and rapport. Speakers would play on people’s tendency to perceive more positively others whom they perceive more similar (including linguistically) to themselves (e.g., Bradac et al., 1988). Moreover, according to CAT, speakers in greater need to gain another's social approval should be more likely to converge (e.g., Giles & Coupland, 1991; Gregory & Webster, 1996).

Within the socio-affective framework, social factors (like the motivation to affiliate or the perception of positive affect) have also been interpreted as variables that modulate the degree
of an on-going tendency of people to match each other’s (also linguistic) behaviour. Importantly, both social routes assume that the tendency to linguistically converge reflect not only the positive but also the negative dynamics between speakers, and speakers would converge less or diverge when wishing or perceiving disfavour and disaffiliation. For instance, people converged less on the accent of a narrator that they perceived as ethnically biased against their own culture (Bourhis & Giles, 1977; see also Doise, et al., 1976).

At the same time, separate evidence suggests that unwelcome interlocutors do not necessarily lead to a reduction in linguistic convergence or divergence. Speakers were found to increase linguistic convergence with difficult interlocutors as a compensatory strategy to ensure smooth interaction in complex social situations (Balcetis & Dale, 2005; cf. compensatory social route, Chapter 2).

3.5 **FINALLY BRINGING THE TWO TOGETHER**

So, how can the intent of deceiving affect a speaker’s tendency to lexically converge with their deceived interlocutor? Recall that we defined deception at the level of interactional goals (i.e., to induce the other to believe something that the speaker know is untrue). Our study focuses on deceivers that tell lies (i.e., affirm as true something that they know is false). Also, we operationalise lexical convergence as in previous interactive priming paradigms (cf. section 2.5.1): as participants’ tendency to reuse the suitable-yet-unusual name for objects introduced by their interlocutor against the (assumed) preference to opt for the most common name.

What does this all mean concretely? Let’s go back to our example of a witness lying to the investigator of a murderer about what she saw that morning in the park. If the liar hears the investigator asking her whether the guy she allegedly saw running away was holding a *pistol*, how would the liar refer to the weapon (given that she’s making up the facts)? Would she reuse the term used by the deceived investigator or go for the most common *gun*? Based on the reviewed evidence, we can hypothesize two families of predictions linked to the two main characteristics of deception (i.e., cognitive/emotional load vs. high-level interactional activity) and the two main families of explanations of linguistic convergence (i.e., automatic-cognitive accounts vs. social accounts).
The cognitive link

It has been suggested that deception and lying are complex cognitive tasks, underpinned by an intricate net of mental operations that must be approached in terms of associated cognitive and emotional loads (e.g., Vrij et al., 2015; Walczyk et al., 2005). There is no direct evidence regarding a possible effect of cognitive and emotional loads on lexical (or more generally, linguistic) convergence (for a similar observation see Weatherholz et al., 2014).

However, at a linguistic level, indirect evidence suggests that extra load may impair speakers’ attentional resources and so interfere with the memory encoding and the related automatic priming mechanisms, thus reducing speakers’ convergence (and alignment) tendency. For instance, in Horton and Gerrig (2005a), participants described pictures to two partners either systematically (e.g., dogs to person A, fish to person B) or unsystematically (e.g., flowers to both A and B). In a second phase they re-described pictures from all categories to both partners individually. When they described a picture that was new for that partner but that they had already named for the other, they were more likely to use appropriate new denotations in the less cognitively demanding condition (the systematic one where the discrete division of categories between partners facilitated recalling which item they could not have named for the current addressee). This means that they showed greater non-adaptive self-consistency in their use of referring expressions under increased memory burden. Consistent with these results, Ivanova, Wardlow, Gollan, and Ferreira (2013) reported weaker priming effects under memory load at a syntactic level.

Nonetheless, evidence from behavioural mimicry (i.e., matching of each other’s unintentional behaviour), indicates a possible opposite effect. Van Leeuwen, van Baaren, Martin, Dijkstra and Bekkering (2009) reported that high cognitive demand (via increased working-memory load) enhanced individuals’ automatic imitative responses (i.e., faster reaction times when copying another’s finger movements). Together, this range of evidence suggests that there are two possible outcomes to the question of lexical convergence in deceptive interaction (e.g., will the deceiving witness re-use “pistol” or opt for “gun”?) within the automatic priming framework.
Cognitive-Less: The increased cognitive load in liars may hinder the depth with which they process others' speech. This would result in a reduced activation of the heard (i.e., primed) terms at the level of their memory representation and, thus, a weaker tendency to re-use those terms (i.e., lexically converge) (cf. Ivanova et al., 2013). In addition, the cognitive inflexibility found in liars suggests that liars could tend to stick to their own ways of calling objects (cf. Horton & Gerrig, 2005a). Recall that, according to IAM, linguistic convergence and alignment are functional to developing shared mental models between interlocutors. Self-consistency in lexical choices may help deceivers succeed with their lies by maintaining two separate mental representations for what they know is reality, and what is falsehood but that they share as reality with their addressee.

Cognitive-More: The reduced attentional and executive-control resources may prevent liars from monitoring the appropriateness of terms in the other's speech (cf. Van Leeuwen et al., 2009). This would result in a higher tendency to imitate, and so lexically converge. In contrast, the lack of cognitive load would allow truth-tellers to monitor for the appropriateness of terms and use what they find is the most appropriate (or common) name for an item (i.e., weaker lexical convergence tendency).

The affective/social link

As we saw, deception and lying are also forms of communication. There are aspects of the interactive process that seem to facilitate them. Increased mentalising and other-monitoring, for instance, have been shown to lead to enhanced sensitivity towards the other's behaviour and to refined manipulations of the other's feedback and beliefs (cf. Burgoon et al., 1998). In addition, the language of deceivers has been found to show attunement with their victims on the one hand (Hancock et al., 2008), but also distance and non-commitment on the other (Zuckermann et al., 1981).

Linguistic convergence has been shown to be used by interlocutors to gain social approval (cf. Natale, 1975) and win over difficult interlocutors (cf. Balcetis & Dale, 2005). In line with this evidence, driven by the desire to be believed and instil trust in their victim, deceivers may make “strategic” use of lexical choices. At the same time, research suggests that people tend to converge less as a response to social situations that are perceived as
negatively charged, probably to increase social distance or as a result of disaffiliation (cf. Bourhis & Giles, 1977): feelings that deceivers may have toward their victim. Therefore, here again the effect might go in two directions.

*Social-Less*: Deceivers may feel the need to distance themselves from their “victim” and their stories. They may reflect this need in linguistic distance (cf. Bourhis & Giles, 1977) and show a weaker tendency to lexically converge than truth-tellers.

*Social-More*: Deceivers, guided by the desire to be believed and avoid creating suspicions in their deceived interlocutor, may feel stronger the need to be liked, install trust in their interlocutor and create the grounds for a smooth interaction (cf. Natale, 1975; Balcetis & Dale, 2005). As a result, they may show a higher lexical convergence tendency than truth-tellers.

In three experiments, we sought to distinguish between the direction of the possible effect of deceiving on lexical convergence (i.e., are liars converging more or less with their deceived partner? Less vs. More comparison), and whether the effect is due to cognitive or social factors (Cognitive vs. Social).

### 3.6 Experiment 1

Experiment 1 is a first exploration of how lying may affect lexical convergence in real-time spoken interaction: are we more or less likely to lexically converge when we lie to our conversational partner?

We created a new paradigm in which pairs of naive participants engaged in a “crime” scene investigation task in the role of Investigator and Respondent. The investigator had to rely on the respondent's knowledge, who witnessed the “crime”, to identify the location of the objects found at the scene so that she could accurately reconstruct it and identify the culprit. We asked half of the Respondents to deceive their partner Investigator by lying about the location of the objects so that the wrong culprit could be blamed. Importantly, the Investigator led the conversation and introduced some of the objects with highly acceptable-yet-dispreferred names (e.g., *pistol* for gun). We were interested in whether Respondents reused their partner's unusual choice of names, under the assumption that participants were
unlikely to use the dispreferred names unless these were previously used by their dialogue partner (lexical convergence).

Our paradigm met the following requirements which we deemed essential to guarantee that deception and lying in our study could lead to similar degree of cognitive and emotional demands, as well as engagement, as their real-life counterparts. We included deception in a real social exchange between naïve people. Importantly, the deceived person was blind to the deceptive manipulation throughout the whole experiment (and none of the participants reported they realised having been deceived in the post-interaction questionnaire and in the debriefing talk with the experimenter). It was clear to deceiving participants what they were lying about and why they were deceiving. Moreover, deceivers could clearly perceive their deception and lies as having clear consequences on the interaction and on the actions/thoughts of the deceived.

3.6.1 Hypotheses

To distinguish between the cognitive vs. social possible hypotheses, we included a within-subjects cognitive load manipulation (i.e., remembering a six-digit string), under the assumption that this would selectively interfere with the linguistic convergence tendency. In our study, we assumed that automatic-cognitive mechanisms (priming) operate independently from social-goals and social mechanisms. We also assumed that cognitive load would interact with the cognitive but not the social mechanisms.

If the difference in lexical convergence between truth-tellers and deceivers is due to cognitive factors that interfere with the underlying automatic priming mechanisms by impairing deceivers' attentional resources (cf. IAM), then we hypothesized that the effect of increasing cognitive load (by imposing another control-demanding task) would be additive, thus interfering with both lying and truth-telling but more so with lying (as already cognitively taking). We hence expect two main effects: one for the deception manipulation, with liars showing weaker or greater lexical convergence than truth-tellers, and one for the cognitive load manipulation, with respondents under high load showing weaker or greater lexical convergence effect than under low load. We had no prediction as far as an interaction is concerned, but it is likely that cognitive load would interfere more with lying than with
truth-telling, and so selectively increase the convergence effect more in the lying high-load condition.

If the lexical convergence effect is driven by affectedness-related reasons or social outcomes (cf. CAT), we should only find a main effect for the deception manipulation but no effect of the load manipulation, with lying respondents showing weaker or stronger convergence than truth-tellers, but no difference between cognitive load conditions.

3.6.2 Method

3.6.2.1 Participants

64 native speakers of British English (11 male; 23 female; mean age = 21.0, SD = 2.5) from the University of Edinburgh's student community, paired into 32 same-gender dyads of unacquainted members, participated in the experiment. They were either paid £6 or received school credits.

3.6.2.2 Material

We constructed 24 experimental items. An experimental item consisted of a prime pictured object, the prime dispreferred name for that object and a target pictured object (identical to the prime pictured object). For example: the prime picture of a gun/pistol, the prime dispreferred name pistol and the target picture of a gun/pistol. We also constructed 48 filler items consisted of a filler prime pictured object, a highly-preferred name for that object and a filler prime pictured object. For example: the prime picture of a cigarette, the name cigarette, and the target picture of a cigarette. Eight of the experimental items were taken from Branigan et al. (2011). The remaining 16 experimental items were chosen from a sample of pre-tested pictures that had been assessed as having a highly preferred name (i.e., preferred by 80% of assessors) and a fully acceptable (i.e., rated at least 5 on a scale from 1 to 7 from a different group of assessors) but dispreferred alternative name (full details about the item selection process are provided in Appendix A).

Items were presented to participants in groups of six (2 experimental items and 4 filler items) as objects that formed part of a scene which also contained three or four characters. Each scene represented a detective case (e.g., who killed the man?) to be solved by the
Investigator. Of the two experimental items in the scene, one had a key role: its correct localisation was essential to understand who did the action and hence to solve the case (*key object*; e.g., the gun/pistol in “who killed the man?”). The other experimental item was not essential to solve the case (*secondary object*; e.g., the glasses/spectacles in “who killed the man?” Figure 1). Investigator and Respondent saw different copies of the same detective scene. Investigators' copies were incomplete (Figure 1): the six objects were missing and were displayed separately on the screen underneath the scene together with a name for each object (i.e., the dispreferred name for the prime item). Within the scene, 8 red dots represented possible locations for the objects (6 real and 2 fake locations). Respondents saw complete copies of the scene with all six objects correctly located (Figure 2). Each scene was combined with the question “Who did X?” (e.g., “*who killed the man?*”). The question was presented to Investigator and Respondent at different times: Respondent saw it during the initial familiarisation phase, whereas Investigator only saw it at the end of the familiarisation phase when the “interrogation” began (see Procedure).

*Figure 1* Example scene “Who killed the man?” as seen by the Investigator
Figure 2 Example scene “Who killed the man?” as seen by the Respondent.

We constructed two lists (A and B), each containing 12 scenes (for a total of 24 scenes), so that each experimental item covered both roles of key and secondary object (i.e., the gun/pistol was a key object in list A but a secondary object in list B) but was only seen once by participants, who were randomly assigned to one of the lists. Each list was divided into two sub-lists (i.e., A1 and A2, B1 and B2); their order of presentation was counterbalanced across participants. All sub-lists occurred an equal number of times in the two Load conditions.

Items and scenes were created by the experimenter with the vector graphics editor InKscape. The experiment was run on E-Prime; the verbal interaction was recorded via a Marantz PMD661 device and two clip-on microphones.

3.6.2.3 Procedure

Pairs of participants were randomly assigned to the Truthful or Deceptive condition. Due to the nature of the experiment, the experimenter was blind to the main experimental manipulation during her initial interaction with the participants and as far as it was possible into the experimental session. This was guaranteed as follows: two participants at a time
came to the lab and were taken to a room by the experimenter. They sat at two tables placed side by side on which two computer screens were placed. The experimenter introduced the study as an investigation into how well people communicate when they cannot see each other, and told them that they were about to engage in a cooperative detective game. One participant was randomly appointed the role of Investigator, the other one the role of Respondent. They then received written information sheets and instructions, and were given 5-7 minutes to read them through. The instructions differed for Investigators and Respondents but, for each role, they were the same in the Truthful/Deceptive conditions. Participants could then ask questions. The experimenter then showed them a practice trial during which participants could see each other's screen and so both copies of a practice scene. After this, participants wore recording microphones and a divider was placed between them. At this point the experimenter left the room and drew lots as to which condition the pair would fall into: participants then received consent forms and the lying participants were also given a message that instructed them to deceive their partner by lying about the location of the objects in the scene. The experimenter collected the forms, left the room and the experiment started. All instructions are available in Appendix B.

Investigator and Respondent saw copies of the 12 detective scenes on separate computer screens. Participants' official goal as a pair was to allow the Investigator to guess correctly who, in each scene, committed the action. To do this, they had to correctly reconstruct the scene by commonly determining the location of the six missing objects. Lying Respondents had to deceive their partner by lying about the location of the objects so as to lead their partner to identify a wrong culprit.

The experiment consisted of two blocks of trials, each containing six scenes (i.e., sub-lists A1 and A2): a high (cognitive) load and a low (cognitive) load block, with the order of blocks and sub-lists counterbalanced across participants. In the high-load block, respondents had to remember a 6-digit code, displayed for 2000ms before the scene. Officially, the code represented the detective-agency file number for that case. The low-load block contained no extra task. The order of scenes within each block/sub-list as well as the order with which the six objects were displayed on the Investigator’s screen was randomized per each participant.
At the end of each trial, participants performed an accuracy test that asked them to recollect where the objects were located in the scene. The aim of the accuracy test was two-fold: to make sure that participants carefully went through the location of all six objects; to attest that the cognitive Load manipulation worked (participants should recall less in the high-load conditions).

An overview of the sequence of play is displayed in Figure 3. Below, we give details of each phase in the game (see also Figure 4).

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Respondent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Trial instructions (i.e., reminder of different phases)</td>
<td>1. Instructions (i.e., reminder of different phases and whether the case is filed (high-load) or non-filed (low-load))</td>
</tr>
<tr>
<td>2. (blank)</td>
<td>2. 6-digit number displayed for 2000ms (high-load) or message that “this case has no file number” (low-load)</td>
</tr>
<tr>
<td>3. 20sec familiarisation phase (no question displayed)</td>
<td>3. 20sec familiarisation phase (question displayed)</td>
</tr>
<tr>
<td>4. Investigation phase (question displayed)</td>
<td>4. Investigation phase (question displayed)</td>
</tr>
<tr>
<td>5. Recap</td>
<td>5. Recap</td>
</tr>
<tr>
<td>6. (blank)</td>
<td>6. Asked to type in the 6-digit code (high-load) or read message “case completed” (low-load)</td>
</tr>
<tr>
<td>7. Type in response to the question “Who did a certain action?”</td>
<td>7. (blank)</td>
</tr>
<tr>
<td>8. “Please wait” message</td>
<td>8. Accuracy test</td>
</tr>
</tbody>
</table>

**Figure 3** Overview of the different phases in a trial.

**Familiarisation phase**

Each trial began with a familiarisation period of 20sec during which participants independently inspected their copy of the scene. During this phase, only the Respondent could see the question introducing the event making up the case (e.g., “who killed the man?”) at the top of the scene. The question appeared at the top of the Investigator’s scene after the 20sec elapsed and the Investigation phase began.
Investigation phase

For each object, the Investigator was instructed to ask a question like "Is [object-X][guessed location]?” (e.g., “is the cigarette in the hand of the sitting man?”). The Respondent was instructed to answer in the form: "No, [the object-X] is [location]" (e.g., “No, the cigarette is in the hand of the woman”), or “Yes, [the object-X] is [location]” (e.g., “Yes, the cigarette is in the hand of the sitting man”).

Recap

The Respondent re-described the locations of all 6 objects in one go. It was during this phase that we measured whether they converged with the lexical choices of the Investigator.

Accuracy test

After the Recap, participants were presented with new copies of the scene. Each copy contained a red symbol and came with the question: "What was actually in the location indicated by the red symbol?". Respondents did this 6 times (one for each item in the scene), whereas naive Investigators did it 8 times (one for each of the 8 possible locations) and typed in “nothing” if no object was in that location.
At the end of the experimental session, participants filled in a questionnaire. To assess whether Investigators suspected having been deceived, we asked them whether, if told that
some participants in the experiment were telling the truth and some were lying, they thought their partner was a truth-teller or a liar. We also checked whether lying Respondents showed a higher degree of consideration for their partner (i.e., whether they had been trying to be particularly kind to their partner) and asked both participants to judge each other’s attitude in the game. Copies of Investigators’ and Respondents’ questionnaires are available in Appendix C.

Participants were eventually debriefed by the experimenter on the real purpose of the study, who also re-checked for suspicions.

3.6.2.4 Design

This study used a 2 x 2 mixed factorial design with Context (Truthful; Deceptive), and cognitive Load (High; Low) as factors. Context (between-subjects, within-items) represents whether the Respondent was deceiving the Investigator. Load (within-subjects; within-items) represents our cognitive load manipulation. The name used by Respondents to refer to the experimental items in the Recap phase (Recap Name) was coded as converged if it was the dispreferred name introduced by their partner for that picture, and non-converged otherwise.

Lexical convergence was measured during the Recap (and not the Investigation) phase because the difficulty of lying and deceiving does not simply rest in producing a response that conflicts with what one knows is true, but in maintaining the consistency of (and faking a belief in) the false narrative. We assumed the associated cognitive and emotional demands would be better present in a Recap phase following the initial Investigation phase.

3.6.3 Analysis

3.6.3.1 Exclusion criteria and analysis specification

The data of 32 Respondents were collected: 16 per Context condition. Each Respondent responded to the full set of 24 experimental items, nested within 12 scenes, resulting in 768 experimental observations (384 per Context condition). However, the data of 6 Respondents (4 Truthful; 2 Deceptive) could not be included because they failed to compile to the Instruction and skipped the Recap phase. This made a total of 624 data points (228 Truthful, 336 Deceptive; 312 per Load condition). We further excluded 58 observations where
Respondents failed to recap the location of that experimental item (16 Truthful, 42 Deceptive) and additional 79 where Respondents introduced the item themselves or Investigators introduced the item using the preferred name for that object instead of the scripted dispreferred.

The final dataset contained 487 Recap Names (241 Truthful, 246 Deceptive; 241 Low Load, 246 High Load). We coded Recap Name so that converged = 1, non-converged = 0. Data were analysed using logit mixed-effects models (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). All models were fitted in R (version 3.2.0; R Core Team, 2015), using the {lme4} package (version 1.1.11; Bates, Maechler, Bolker, Walker (2015). Whenever possible the maximal random structure justified by the design was included (Barr, Levy, Scheepers, & Tily, 2013); in case of failed convergence, we removed the correlations between the random effects, first by items and then by subjects. In case of perfect random correlations (|r| = 1), we removed them (Appendix to Barr et al., 2013) and made sure the removal was justified by likelihood ratio $\chi^2$ tests comparing the model with and without the random correlations (Bates, Kliegl, Vasishth, & Baayen, 2015).

As fixed effects, we included the two deviation-contrast coded predictors Context (Truthful = .5; Deceptive = -.5) and Load (High = .5; Low = -.5), and their two-way interaction. The model also included by-subjects and by-items random intercept, as well as by-subjects random slope for Load and by-items random slopes for Load, Context and their interaction. The significance of the fixed-effects parameters was evaluated via a Wald's Z statistic, which assesses whether the regression coefficients are significantly different from zero given the estimated SE.

### 3.6.4 Results

Overall, respondents converged with the dispreferred name introduced by their partner on 75% of occasions. In both Context and Load conditions, respondents produced almost identical proportions of converged names (Truthful: 75%, Deceptive = 75%; Low = 75%)

---

11 At first, we also included a by-scene random intercept but as it did not improve model fit as shown by likelihood ratio test results, also on the ground of the small amount of degrees of freedom in our data, we did not include it in the final model. This choice did not affect the significance of the fixed-effects predictors.
High = 76%). Figure 5 shows the observed proportions of converged Recap Names by Load and Context, together with the mixed logit model predictions. Frequencies are displayed in Table 1.

### Table 1

**Frequency of converged (non-converged) Recap Names by Respondents per Context and Load condition**

<table>
<thead>
<tr>
<th>Load</th>
<th>Truthful</th>
<th></th>
<th>Deceptive</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>key objects</td>
<td>secondary objects</td>
<td>key objects</td>
<td>secondary objects</td>
</tr>
<tr>
<td>low</td>
<td>46 (12)</td>
<td>42 (14)</td>
<td>53 (10)</td>
<td>43 (21)</td>
</tr>
<tr>
<td>high</td>
<td>50 (16)</td>
<td>48 (13)</td>
<td>46 (12)</td>
<td>43 (18)</td>
</tr>
</tbody>
</table>

*Figure 5* Observed mean proportions of converged Recap Names in the Truthful and Deceptive conditions by Low Load (red circle) and High Load (blue triangles). Error bars represent 95% bootstrapped confidence intervals based on the subject-wise condition means. Predicted mean proportions from the mixed-effects logit model (stars) and horizontally jittered by-subjects observed proportions are also displayed.
Contrary to all our hypotheses, neither Context, Load or their interaction significantly predicted lexical convergence ($p > .4$). Respondents were as much as likely to converge on the dispreferred name introduced by their partner Investigator under all our experimental manipulations (model summary in Table 2).

### Table 2
Summary of the mixed logit models for Experiment 1 ($n = 487$, dyad = 26, item = 24)\(^{(A)}\)

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Wald's test</th>
<th>$\beta$</th>
<th>SE</th>
<th>odds-ratio</th>
<th>95% profile CI</th>
<th>Z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>1.38</td>
<td>0.28</td>
<td>3.97</td>
<td>[2.29; 7.26]</td>
<td>4.92</td>
<td></td>
</tr>
<tr>
<td>Load</td>
<td></td>
<td>-0.18</td>
<td>0.24</td>
<td>0.83</td>
<td>[0.51; 1.35]</td>
<td>-0.76</td>
<td>0.448</td>
</tr>
<tr>
<td>Context</td>
<td></td>
<td>-0.16</td>
<td>0.35</td>
<td>0.85</td>
<td>[0.41; 1.76]</td>
<td>-0.46</td>
<td>0.647</td>
</tr>
<tr>
<td>Load:Context</td>
<td></td>
<td>-0.23</td>
<td>0.48</td>
<td>0.80</td>
<td>[0.30; 2.10]</td>
<td>-0.47</td>
<td>0.647</td>
</tr>
</tbody>
</table>

Random effects (Std Dev)
- Subj: Intercept: .66
- Subj: Load: < .00
- Item: Intercept: 1.03
- Item: Load: < .00
- Item: Context: < .00
- Item: Load: Cont: < .00

\(^{(A)}\) We forced independence between all random effects because of failed convergence. All correlations between fixed effects $< |.01|$

3.6.5 Exploratory Analysis

**Object Role**

Deceiving had no overall effect on speakers' lexical convergence tendency. However, it may be that respondents showed a different likelihood to re-use the dispreferred name of objects depending on whether the object was key or not to solving the detective case and that this selectively interacted with whether they were deceiving their partner or telling the truth. That is, the motivation to succeed in the deception may have been stronger for key objects; this higher desire may have increased liars' tendency to lexically converge for key objects.
because they were functional to their interactional goal. To test this, we constructed an additional mixed logit model with sum-contrast coded\textsuperscript{12} Context (Truthful = -1; Deceptive = 1), Object Role (Key = -1; Secondary = 1) and their interaction as predictors. No effect was found for Object Role ($\beta = -0.21$, SE = 0.13, $p =.116$, odds-ratio = 0.81, 95% profile CI [0.61; 1.07]); lying and truth-telling respondents showed similar lexical convergence tendency when referring to key or secondary objects ($\beta = -0.20$, SE = 0.14, $p =.153$, ratio of odds-ratio = 0.82, 95% profile CI [0.59; 1.10]).

\textit{About the high lexical convergence rate}

A striking aspect of the collected data was the very high rate of lexical convergence across all conditions. The average by-Respondent convergence rate in our experiment was 75\% (range: 38\% - 96\%), well above the reported average rate of 43\% and 58\% in Branigan et al. (2011) who used a similar name-alternative paradigm. This particularly high convergence rate was confirmed also by only looking at the eight items that were the same as in Branigan et al. (2011) (average convergence: 71\%, range: 17\% - 100\%). We could think of two reasons that could explain this particularly high convergence effect. The first one is linked to the accuracy test. The fact of having to type in objects' name may have led participants to pay particular attention to names and infer that the way objects were called was important for the game. To rule out this possibility, we checked the convergence effect in the first scene for each respondent, where an effect of accuracy test couldn't yet be present. The analysis showed an average convergence rate of 77\% (Truthful: 83\% ± 8\% SE; Deceptive: 69\% ± 12\% SE), which was not different from the overall tendency, contrary to this hypothesis.

A second possibility is that once the experimental items were placed within the context of a scene, participants stopped preferring the preferred name (e.g. “gun” over “pistol”). We tested this by asking 14 new participants to spontaneously name all experimental items in the context of their scene. Each item was only named once by each person. We stuck to the threshold used in the pre-test and classified as preferred the name spontaneously used by at least 80\% of participants. Only 11 out of the 24 original preferred names could still be

\textsuperscript{12} We went for sum contrast coding as it facilitated model convergence. The model included by-subjects random intercept and slope for Object Role, and by-items random intercept, and slopes for Context, Object Role and their interaction. We forced independence between all random effects. All correlations between fixed effects <|.07|. 

93
classified as *preferred* in the post-test (mean % preference: overall 73%, confirmed preferred 96%, non-confirmed preferred 53%). This suggests that the original preferred-dispreferred name classification was likely no longer in hold once the objects were placed in the context of a scene.

### 3.6.6 Discussion

We did not find any effect of deception or cognitive load on the extent with which participants converged with the unusual lexical choices of their interlocutor. Why could this be?

The least attractive explanation is that this is a real null effect: deceiving does not affect the degree with which we lexically converge with our victim. However, exploratory analyses suggest that this may not be the whole true story (no pun intended). Participants, in fact, showed extremely high degrees of lexical convergence across all experimental conditions; what was this due to? As observed in the post-test, it is possible that the preferred names stopped being preferred once in the context of the scene and respondents converged to such a great extent on the dispreferred name because that's the way they would have referred to that object anyway. Another possibility links the high convergence rate to an excessive cognitive load. Respondents were not given the solution to the case but had to quickly figure out the culprit by themselves, and it is plausible that this required most of their cognitive resources. The initial familiarisation phase of 20 sec was probably not long enough to understand who the culprit was and be ready to answer the Investigator's questions (in either Context). This might have added extra cognitive load to Respondents’ attentional resources which, impaired, prevented them from monitoring the appropriateness of terms introduced by the Investigator, thus causing the massive convergence effect (via the *Cognitive-More* link), and cancelling out any possible effects of our experimental manipulations. We addressed both issues in Experiment 2.

### 3.7 Experiment 2

Following the uniformly high levels of convergence obtained in Experiment 1, we decided to simplify the experimental paradigm and have participants engage in a “classic” picture-naming/-matching task. At the same time, we took a theoretical step backwards and decided
to test only for the direction of the potential effect of deception on lexical convergence, thus overlooking for the moment whether this may be due the underlying social vs. cognitive mechanisms.

So in Experiment 2, participants and a confederate took turns to name pictures for their partner and choose pictures that matched their partner's name. We used the same logic as in Experiment 1: for some of the pictures, participants were primed by the confederate with highly acceptable-yet-dispreferred names (e.g., boombox for stereo). This time, however, the participant named the same critical picture always two filler turns later. As before, we were interested in whether they picked up and reused their partner's unusual choice of name against the tendency to use the most common name for that picture (lexical convergence).

Every few trials, participant and confederate did a memory test to identify pictures not selected in their previous matching turns; to perform well on the test, the correct pictures had to be previously named by their partner during the participant’s matching turns (see section 3.7.1.4). We manipulated the intentional context of the interaction and asked half of the participants to deceive their partner by sometimes naming the wrong object with the aim to hamper their partner's performance on the memory test (with the incentive of financial reward if they did so). This allowed us to examine whether participants' tendency to lexically converge with a conversational partner was selectively affected by whether they were being deceitful (vs. truthful) towards them. The aim of the memory test was two-fold. First, it measured the extent to which deception taxed participants' cognitive resources; if deceiving is indeed cognitively demanding then deceivers’ performance on the test should be worse compared to truth-tellers. Second, it provided participants in the Deceptive condition with a believable reason for lying that had tangible consequences on their partner. Importantly and in contrast to Experiment 1, the memory test was not used to selectively affect convergence and took place in all blocks.
3.7.1 Methods

3.7.1.1 Participants

48 native speakers of British English aged between 18-23 (12 male; 36 female; mean age = 19.8, SD = 1.7) from the University of Edinburgh's student community participated in the experiment. They were paid £3.5 or received school credits.

3.7.1.2 Material

We created 28 experimental items using the exactly same procedure as in Experiment 1 (see complete list in Appendix D). Experimental items were divided into two lists so that, in the Deceptive condition, participants lied about items from one list but not about items from the other (counterbalanced across participants). We also constructed 162 filler items which were not subject to any constraints apart from being easily recognizable. All items were 200x220px large. The experiment was run on E-Prime; the verbal interaction was recorded via a Marantz PMD661 device and two clip-on microphones.

A post-interaction questionnaire (see Appendix E) inquired about participants' perception of the names used by their partner, and whether, for any object, they felt using different names than the ones used by their partner (if so, why).

3.7.1.3 Design

The study used a between-subjects design with participants induced to either being truthful or deceiving their partner (i.e., a confederate naive to the main manipulation). Context (Truthful vs. Deceptive) was our main independent variable. In each condition, participants performed 7 blocks of trials, each consisting of 12 matching and 12 naming turns in alternation, followed by a memory test. In the memory test, participants identified, among a matrix of 16 objects, the 4 objects that they had only seen but not selected so far on their previous matching turns.

In each block, 4 images were experimental items. The experimental item was always named first by the confederate using its dispreferred name (e.g., boombox) (critical matching turn). After two intervening filler turns (a participant's naming and matching turn), it was the
participant's turn to name that the same experimental item \((critical\ naming\ turn)\): the name they used to refer to it was coded as \(\text{converged}\) if this was the dispreferred name previously used by the confederate, and \(\text{non-converged}\) otherwise. Target Name (converged vs. non-converged) was our binomial dependent variable.

In the Deceptive condition, participants named half of the experimental items while lying (2 per block, 12 in total) and the other half while telling the truth (2 per block, 12 in total). Case (Truth vs. Lie) was the independent variable representing this aspect of the design. This extra manipulation was introduced to reflect the fact that when people try to deceive someone they do not lie about everything they say. We were interested in pinning down whether the lexical convergence tendency was only affected by the fact of deceiving or, more specifically, by lying within that deceptive context.

An accuracy measure was estimated for each memory test as a binomial variable; it was coded as \(\text{correct}\) if all four previously-unselected objects were identified, \(\text{incorrect}\) otherwise. Block was a discrete numeric covariate (range 1-7) representing time.

### 3.7.1.4 Procedure

Participant and confederate arrived to the lab and were welcomed together. The confederate was a 24 year old female student from London. They were taken to a room by the experimenter and sat at two tables, placed side by side and separated by a partition. The confederate always sat on the right table. On each table there was a computer screen, a keyboard and a mouse. The experimenter introduced the study as an investigation into the effectiveness of collaboration between people that cannot see each other, and told them that they were about to engage in a joint picture naming task. The experimenter then explained the structure of the different turns (i.e., matching, naming, memory) to the participants and emphasised that in order to do well on the memory test the right objects had to be named by their partner and matched accordingly. Participants were also encouraged to be fast and accurate. At this point participants read the information sheets outlining their rights as research subjects. The experimenter, who had so far been blind to the experimental manipulation, took this time to randomly assign the participant to the Deceptive or Truthful condition and a trial list. Participants were then given appropriate written instructions and
consent forms to sign (Appendix F). The instructions for the Deceptive participants contained a private message instructing them to deceive their partner. Their goal was to hinder their partner's performance on the memory test; they could achieve this by deceiving their partner and naming the wrong object when this was possible: when they saw the word 'LIE' after the fixation cross. They were also assured that their partner would never lie to them. In contrast, participants in the Truthful condition had the goal to allow each other to select the right pictures so that they could each accurately complete their memory test.

After the experimenter collected the forms, participants started a practice session consisting in a normal block of naming and matching turns, and a memory test, that only contained filler items. After this practice, if participants had no questions, the experiment started.

Participants' critical matching turns

Every experimental session started with the confederate naming the object and the participant matching the object (Figure 6). On participant's critical matching turns, a matching and a mismatching picture (always a filler) appeared side by side on the participant's screen. The matching picture was an experimental item (Prime Picture) (e.g., stereo/boombox). Only the matching (prime) picture appeared on the confederate's screen, followed, after 1000ms, by the text “It's now your turn to name” together with the dispreferred name for that picture (Prime Name) (e.g., boombox). The confederate named the picture and the participants used the left/right arrow keys to select the matching picture on their screen. The picture selected got highlighted and the choice remained visible for 1000ms. Matching experimental pictures appeared an equal number of times on the left and right side of participants' screen. Critical matching turns were the same in both Context conditions.

Participants' filler matching turns

Participants' filler matching turns were identical to the critical matching turns with the exception that both matching and mismatching pictures were filler items. The confederate was not prompted with a name for the filler object.
Participants' critical naming turns

Participant's critical naming turns differed according to Context and Case conditions. On the critical naming turns in the Truthful Context and in the Truth Case in the Deceptive Context conditions (Figure 6), after a fixation cross (700ms), the Target Picture (e.g., stereo/boombox) and a distractor picture (e.g., club) appeared side by side on the participant's screen. The side on which the Target Picture appeared was randomized. The Target Picture also appeared on the confederate's screen together with a different mismatching picture (e.g., chicken). After 1000ms, a yellow box surrounded the Target Picture on the participant's screen and the text “it's now your turn to name” appeared. The participant named the highlighted Target Picture. The confederate selected the matching picture on her screen.

On the critical naming turns in the Lie Case in the Deceptive condition (Figure 7), after the fixation cross (200ms), participants saw the message 'LIE' (500ms); this was followed by the Target Picture (e.g., stereo/boombox) and a distractor (e.g., club) appearing on the screen. The exact same two pictures appeared on the confederate's screen. After 1000ms, a blue box surrounded the distractor on the participant's screen together with the text “it's now your turn to name”. The participant, prompted to lie by the message, named the non-highlighted object (i.e., the Target Picture).

Hence, critical naming turns in the Deceptive Context Lie Case condition differed from the critical naming turns in the Truthful Context and Deceptive Context Truth Case condition in that the Target Picture was the one that did not get highlighted. In both cases, the name used by the participant to refer to the Target Picture (Target Name) was coded for lexical convergence.

In all critical naming turns, the name of the distractor picture was phonologically unrelated to either the preferred or dispreferred name of the Target Picture (i.e., it was not a homophone with either name, and shared no word-initial or word-final segment with either).
Figure 6 Example of critical turns in the Truthful and the Deceptive Context Truth Case conditions. (a) The confederate named the Prime Picture using the dispreferred name (e.g., boombox). After two filler turns, (b) the participant named the same experimental object (Target Picture). Convergence occurred if the participant used the same dispreferred name as the experimenter (i.e., boombox).
Figure 7 Example of critical turns in the Deceptive Context Lie Case condition. (a) The confederate named the Prime Picture using the dispreferred name (e.g., boombox). After two filler turns, (b) the participant was cued by the “LIE” message to name the same but non-highlighted experimental object (Target Picture). Convergence occurred if the participant used the same dispreferred name as the experimenter (i.e., boombox).
Participants' filler naming turns

Participant's filler naming turns were the same as the critical naming turns in the Truthful Context condition with the difference that both matching and distractor pictures were filler items. In each block in the Deceptive condition, there was one Lie filler naming turn. This had the exact same structure as the critical naming turn in the Deceptive Context Lie Case, except that all pictures were filler items. This type of naming turn was included to prevent participants from realizing that each time they were cued to deceive, they had to name a picture which had been named in an unusual way by their partner.

Memory tests

After 12 naming and matching turns participant and confederate performed individual memory tests. The participant's memory test was the same in the two Context conditions. A grid of 4x4 pictures was displayed. 12 were 'old' pictures that their partner (the confederate) had named for them during the participant's previous matching turns (and so, that they should have selected); 4 were 'novel' pictures that they had seen as distractors but shouldn't have so far selected. Participants' goal was to identify the four novel pictures and click on them with the mouse, having only four shots available. At the end of the memory test, a message informed them whether they got all four pictures right (they were not told how many they got wrong if so). Importantly, the memory test was linked to the pictures participants had seen and selected in their matching turns. As such, it was not directly affected by the deceptive manipulation which directly affected participants' behaviour in naming turns only.

The crucial aspect of the memory test is that what count as 'old' pictures are the pictures that were highlighted to be named by the partner; that is, if the partner named the highlighted picture (e.g., heart) but the person accidentally selected the other picture (e.g., button), it was still the highlighted pictures (i.e., the heart) that counted as 'old' picture for the memory test. This is important to understand how, in the participants' mind, deceiving hampered the confederate's performance on the memory test. In the Deceptive condition, the fact that the participant had named three times the 'wrong' non-highlighted object implied that the confederate would recognize as 'novel' (i.e., previously unselected) pictures that were
actually 'old' (i.e., pictures that should have been selected), and as 'old' (i.e., selected) pictures that were actually 'novel' (i.e., shouldn't have been selected), thus hampering her performance on the test. To avoid creating suspicion in the confederate concerning the main manipulation, she was not told about her performance on the memory test. Nevertheless, the confederate was asked to report, after every experimental session, whether she had any comments on the interaction. This would have allowed us to identify the moment in which she would become aware of the Truthful/Deceptive manipulation.

To summarise, in both Context conditions, participants performed 7 blocks of trials and responded to a total of 84 pictures in their naming turns (56 filler and 28 critical items); they selected the same 28 critical items and 56 additional filler items in their matching turns. In the Deceptive condition, the ratio between truths and lies was 3:1 (i.e., two Lie Case critical naming turn, one Lie filler naming turn per block). The participant completed the memory test first. There was always one participant's filler naming turn and one filler naming turn between the Prime Picture being named by the confederate and the Target Picture being named by the confederate. At the end of the experiment, the participant was fully debriefed about the real purpose of the experiment and that their partner was an experimenter's confederate.

### 3.7.2 Results

**Lexical Convergence**

*Exclusion criteria and analysis specification*

The data of 48 participants were collected, 24 per Context condition. The data of two participants in the Truthful condition had to be dismissed before analysis due to age (> 35) and language criteria (i.e., not a BrEN native speaker). There were 23 cases of missing values (1.79%): 2 because participants (Truthful) failed to provide a name for the object, and 21 because participants (Deceptive Lie) failed to lie and so did not name the experimental item. The analysis was based on 1265 observations (Truthful: 614; Deceptive: 651).

We coded Target Name so that converged = 1, non-converged = 0. A mixed logit model (Jaeger, 2008) was conducted to predict whether participants' likelihood to converge with the
dispreferred name introduced by their partner depended on whether they were deceiving her (Truthful vs. Deceptive Context comparison), and, if they were deceiving her, on whether they were lying about that particular item (Lie vs. Truth Case comparison within the Deceptive Context). We followed the same analytical procedure as in Experiment 1 unless otherwise specified. We used Helmert-contrast coding to compare first the Truthful vs. Deceptive Context conditions, and then the Lie vs. Truth cases within the Deceptive context. The model included by-subjects and by-items random intercept, and by-items random slopes for both contrast comparisons\(^{13}\). Overall, participants converged with the dispreferred name introduced by their partner on 57% of times. Table 3 and Figure 8 show observed frequencies and proportions of converged responses by condition.

**Table 3**
Frequency of converged (non-converged) Target Names per Context condition and Case

<table>
<thead>
<tr>
<th>Context</th>
<th>Truthful</th>
<th>Deceptive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lie Case</td>
<td>Truth Case</td>
</tr>
<tr>
<td>Truthful</td>
<td>378 (236)</td>
<td>163 (155)</td>
</tr>
<tr>
<td>Deceptive</td>
<td>179 (154)</td>
<td></td>
</tr>
</tbody>
</table>

\(^{13}\) Context comparison (Truthful = 2; Deceptive Lie = -1; Deceptive Truth = -1), Case comparison (Truthful = 0; Deceptive Lie = 1; Deceptive Truth = -1). All correlations between fixed effects < |.3|. Model df: 1265 N; 46 subjects, 28 items.
Relative to the truth-telling participants, participants deceiving their partner converged significantly less on the dispreferred name introduced by their partner ($\beta = 0.19$, SE = 0.09, $p = 0.038$, odds-ratio = 1.21, 95% profile CI [1.01; 1.45]). As for participants in the Deceptive condition alone, we found no difference in their lexical convergence tendency when they were lying about that object compared to when they were not ($\beta = -0.06$, SE = 0.10, $p = 0.581$, odds-ratio = 0.95, 95% profile CI [0.78; 1.18]). This suggests that once in a deceptive role, participants' weaker convergence tendency was not affected by whether they were lying about that particular object.

Post-hoc, we added Block as a numeric covariate (centred on its mean) to the fixed effects of the model together with its interaction with the two contrast-comparisons in order to determine the time course of lexical convergence in the different experimental conditions\footnote{The model included by-subject random intercept and slope for Block, and by-items random intercept and slopes for all fixed-effects parameters. Model df: 1265 N, 46 subjects, 28 items. All correlations between fixed effects < [0.3].} (see Figure 9).
Consistent with our main analysis, the Truthful vs. Deceptive Context comparison was still significant ($\beta = 0.20$, $SE = 0.09$, $p = .032$, odds-ratio = 1.23, 95% profile CI [1.02; 1.47]). The interaction between Block and the Truthful vs. Deceptive comparison was also significant ($\beta = 0.06$, $SE = 0.03$, $p = .032$, ratio of odds-ratio = 1.06, 95% profile CI [1.01; 1.12]); the positive interaction indicates that, across blocks, lexical convergence rate evolved in opposite directions for the two Context conditions, with deceivers tending to converge less and truth-tellers tending to converge more. Simple effect analysis confirmed the negative effect of Block on lexical convergence in Deceptive Context ($\beta = -0.10$, $SE = 0.05$, $p = .048$, odds-ratio = 0.91, 95% profile CI [0.82; 1.00]), but the non-significance of Block in the Truthful Context ($\beta = 0.06$, $SE = 0.07$, $p = .437$, odds-ratio = 1.06, 95% profile CI [0.91; 1.24]). No other effect was significant ($ps > .1$).

Figure 9 Observed proportions of converged Target Names across blocks in the Deceptive (red circle) and Truthful (blue triangle) condition. Lines display the predicted values from the GLMM regression (grey bands are prediction standard errors).
Cognitive Load (memory test)

The experimental set-up resulted in a total of 322 accuracy observations (Truthful: 154; Deceptive: 168): 7 observations per participant, one for each memory test at the end of each block. Table 4 shows average accuracy rates. We coded Accuracy so that correct = 1, incorrect = 0, and used a mixed logit model with Context as fixed-effects predictor (Truthful = 0), and by-subjects random intercept and by-blocks random intercept and slope for Context.

Table 4

<table>
<thead>
<tr>
<th>Context</th>
<th>Truthful</th>
<th>Deceptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>65 (89)</td>
<td>28 (140)</td>
</tr>
<tr>
<td>%</td>
<td>42.2% ± 10.5%</td>
<td>16.7% ± 7.6%</td>
</tr>
</tbody>
</table>

Context was a significant predictor of accuracy, with deceiving participants less likely to perform correctly on the memory test ($\beta = -1.46$, SE = 0.38, $p < .001$, odds-ratio = 0.23, 95% profile CI [0.10; 0.48]).

Post-hoc, we tested whether the extent to which deceiving participants had lexically converged with their partner could predict their performance on the memory tests. If the weaker lexical convergence was a sign of the greater cognitive effort induced by deceiving, we should find that the less deceivers converged, the worse they performed on the memory test. We recoded Target Name as a numeric variable using partial credit unit scoring: each participant was assigned a value between [0, .25, 0.5, 0.75, 1] for each Block based on the proportion of converged Target Names they had produced in that block (variable Credit Target). We then built a mixed-effects logit model on deceivers’ Accuracy scores with Credit Target as fixed-effects, and by-subjects and by-blocks random intercept. The magnitude of lexical convergence did not predict accuracy performance ($\beta = -0.16$, SE = 0.89, $p = .856$, odds-ratio = 0.85, 95% profile CI [0.14; 4.74]).
3.7.3 Discussion

Our results suggest that being deceptive disrupts the interactive process of lexical convergence. Deceivers were less likely to lexically converge than truth-tellers and converged increasingly less over the course of the interaction, whereas truth-tellers’ degree of convergence remained stable over time.

These results are compatible with the cognitive-less hypotheses of the deception-convergence link, whereby reduced attentional resources led deceivers to shallowly process their partner's speech and so impaired their ability to engage in linguistic coordination (presumably via automatic mechanisms). At the same time, these results do not exclude the social-less hypothesis, whereby deceivers’ “linguistic distance” (reduced convergence) may be a reflection of their sense of estrangement from their victim. The experiment was not set up to distinguish between these two possibilities. However, some insights come from the memory test results. Deceivers performed worse on the memory test, confirming that being deceptive was cognitively more taxing than being truthful, thus corroborating cognitive-load based explanations.

Note that lexical convergence was not modulated by whether deceiving participants were lying (vs. being truthful) about a specific object. This is in line with previous studies that found no behavioural or physiological differences between lies and truths told within a deceptive context (e.g., Carrion et al., 2010; yet see Loy, Rohde, & Corley, 2016). One possibility is that this is caused by a general source of cognitive load, like the one induced by the stress of deceiving (i.e., the emotional pressure of being transgressing the expected pragmatic norm of truth) which hampered deceivers’ attentional resources throughout the whole experiment. However, an alternative explanation deserves our attention. It could be that deceivers’ general attentional resources were impaired by the task-switching but not deceptive nature of their task. That is, deceivers’ attentional resources may have been reduced by having to sometimes inhibit a prominent response (the highlighted card) and produce a conflicting response, but not by inhibiting a true response and producing a conflicting false response. We will address this possibility in Experiment 3.
3.8  **EXPERIMENT 3**

In Experiment 2, deceivers showed a lower tendency to lexically converge than truth-tellers, and also showed poor performance in the memory task. In Experiment 3 we set to investigate what may have reduced deceivers’ attentional resources, a plausible cause of these effects. So, in Experiment 3 we introduced an additional condition in which participants had to sometimes name the non-highlighted object but without deceptive intent (Switch truthful). We measured participants’ tendency to lexically converge, naming latencies, memory performance and self-report feeling of closeness toward their partner.

One possibility is that deceivers converged less than truth-tellers because of the emotional load associated with their deceptive intent of inhibiting a *truth* bias and producing a conflicting *false* response that hampered their conversational partner. If so, then participants in the Deceptive condition should show different patterns of results than participants in both the Switch and Truthful conditions: weaker lexical convergence, slower naming latencies and poorer memory performance. The alternative explanation blames the task-switching nature of the task: the cognitive cost so detrimental for lexical convergence was due to inhibiting a *prominent* response (the highlighted picture) and producing a conflicting response. If so, then participants in the Deceptive and Switch conditions should show same patterns of results: weaker convergence, slower naming latencies and poorer memory performance than participants in the Truthful condition. Both these possibilities rely on the *cognitive-less* link between lexical convergence and deception.

We also tested for a possible (probably additive) role of the *social-less* hypothesis, whereby deceivers’ sense of estrangement from their partner selectively added to the decreased convergence tendency of deceivers. If so, then participants' feeling of closeness should be weaker for deceivers and lead to weaker lexical convergence. An overview of the three possible explanations and their predictions is provided in Figure 10.
3.8.1 Methods

3.8.1.1 Participants

48 native speakers of British English aged 18-23 (15 male; 33 female; mean age = 20.7, SD = 2.9) from the University of Edinburgh's student community participated in the experiment. They were paid £3.5 or received school credits.

3.8.1.2 Design, Material and Procedure

Design, task and procedure were identical to those in Experiment 2 with the following exceptions. The experiment consisted of 9 blocks of trials, with 32 experimental items (four in each block 2-8). We used two confederates (female, 22 and 24 years old, from Southern England). Participants were randomly assigned to one of three Context conditions: Truthful, Deceptive, Task-Switching Truthful (Switch). Each confederate interacted with half of the participants in each condition. The Truthful and Deceptive conditions were identical to Experiment 2. Instructions read by participants in the Switch condition are available in Appendix G.

---

**Figure 10** Overview of the possible causes of deceivers' impaired attentional resources (i.e., task-switching cognitive-load and deceptive-intent emotional load) and predictions for Experiment 3.
As a quick reminder, in the Deceptive condition, participants were cued to deceive their partner by sometimes naming the wrong non-highlighted object for her. They did it every time they saw the word “LIE” after the fixation cross. Participants lied in half of their critical naming turns. In the Switch condition, participants were also cued to sometimes name the non-highlighted object. As in the Deceptive condition, they did so only in half of the critical naming turns. In contrast to the Deceptive condition, however, this had no negative repercussion on their partner’s performance on the memory test and was not linked to any deceptive intent. To the contrary, it was emphasized to participants that naming the non-highlighted picture when they were cued with the word “OTHER” was essential for their partner to do well on her memory test. Case (Other vs. Highlighted) was the binomial variable reflecting this aspect of the design. In the Truthful condition, participants always named the highlighted picture for their partner.

At the end of the experiment, participants were asked to report how close they had felt to their interaction partner by choosing one from a series of seven progressively overlapping circles (cf. IOS scale, Aron, Aron, & Smollan, 1992; see Appendix H).

The Deceptive vs. Truthful comparison was a direct replication of Experiment 2.

3.8.2 Results

For all analyses we followed the same analytical procedures as in Experiment 2 unless otherwise specified. In all analyses, we controlled for confederate’s identity by including Confederate (deviation-coded as confederate1/confederate2 -1/1) and its interaction with Context as fixed effects. As neither Confederate or its interaction were ever significant (all ps > .2), we report the models without them.

Lexical Convergence

The data of 48 participants were collected, 16 per Context condition. There were 21 cases (1.37%) of missing values. The analysis was based on 1515 observations (Truthful: 509; Deceptive: 507; Switch: 499).

Note that in the Deceptive Case conditions Other = Lie and Highlighted = Truth.
Overall, participants converged on the dispreferred name introduced by their partner on 56% of occasions, in line with Experiment 2 (see Table 5).

### Table 5

Frequency of converged (non-converged) Target Names per Context and Case condition\(^{(A)}\)

<table>
<thead>
<tr>
<th>Context</th>
<th>Truthful</th>
<th>Deceptive</th>
<th>Switch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Other</td>
<td>Highlighted</td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>(Lie)</td>
<td>(Truth)</td>
<td></td>
</tr>
<tr>
<td><strong>Truthful</strong></td>
<td>272 (237)</td>
<td>128 (123)</td>
<td>137 (119)</td>
</tr>
<tr>
<td><strong>Deceptive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Switch</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{(A)}\) In both DECEPTIVE OTHER (LIE) and SWITCH OTHER, participants named the non-highlighted picture for their partner but with opposite intents: in the DECEPTIVE LIE they knew this would impair their partner's performance on the memory test, in the SWITCH OTHER condition they knew this would allow their partner to do well on their memory test.

We coded Target Name so that converged = 1, non-converged = 0. To test the source of deceivers' impaired attentional resources (i.e., task-switching cognitive-load and deceptive-intent emotional load), we conducted a mixed logit model on participants' likelihood to lexically converge across the three Context conditions. We used treatment-contrast coding to compare the Deceptive condition (intercept) against the Truthful and the Switch conditions separately. The model included by-subjects and by-items random intercept, and by-items random slopes for both contrast-comparisons\(^{16}\). We compared the Truthful and Switch conditions using the \texttt{gclt} (generalized linear hypothesis test) function from the \{multcomp\} package (version 1.4.4, Hothorn, Bretz, & Westfall, 2008)\(^{17}\). Figure 11 shows the proportion of converged Target names by Context. In contrast to both our hypotheses, deceivers did not produce fewer converged Target names compared to truth-tellers (\(\beta = 0.09, SE = 0.36, p = 0.806, \text{odds-ratio} = 1.10, 95\% \text{profile CI [0.53; 2.28]}\)). There was no significant difference between deceivers and participants in the Switch condition (\(\beta = 0.56, SE = 0.37, p = 0.135,\)

\(^{16}\) Because of failed convergence, we forced independence between the by-items random effects. This was achieved by manually coding the contrast matrix as two numeric vectors. Model df: 1515 N, 48 subjects, 32 items. All correlations between fixed effects < 0.6.

\(^{17}\) Truthful = 1, Switch = -1.
odds-ratio = 1.75, 95% profile CI [0.83; 3.74]), nor between participants in the Switch and Truthful conditions ($\beta = -0.47, \text{SE} = 0.38, p = .211, \text{odds-ratio} = .63, 95\% \text{ CI} [0.30; 1.31])

Despite the lack of significant results in our first analysis, we went on to investigate whether participants were more or less likely to converge on the dispreferred name produced by their partner when they named non-highlighted pictures compared to highlighted pictures, and whether this was further modulated by whether naming the non-highlighted picture was done with a deceptive intent (Deceptive) or not (Switch). We constructed a model with Context (Deceptive = .5; Switch = -.5), Case (Other = .5; Highlighted = -.5) and their interaction as fixed effects, restricting the analysis to the Deceptive and Switch conditions. We included by-subjects random intercept and slopes for Case, and by-items random intercept and slopes for all fixed-effects parameters. Proportions of converged Target names and model predictions are displayed in Figure 12. Neither Case nor its interaction with Context were significant: participants' likelihood of lexically converging was not affected by whether they

[Figure 11: Observed mean proportions of converged Target Names in the three Context conditions. Error bars represents 95% bootstrapped confidence intervals. Predicted mean proportions from the mixed-effects logit model (yellow stars) and horizontally-jittered by-subjects observed proportions are also displayed.]
were naming the highlighted or non-highlighted picture \((\beta = 0.18, \ SE = 0.16, \ p =.240, \ \text{odds-ratio} = 1.20, \ 95\% \ \text{profile CI} \ [0.89; \ 1.64])\), and this was not further affected by whether they were doing it to deceive or instead benefit their partner \((\beta = -0.44, \ SE = 0.30, \ p =.148, \ \text{ratio of odds-ratio} = 0.65, \ 95\% \ \text{profile CI} \ [0.36; \ 1.16])\).

**Figure 12** Observed mean proportions of converged responses when naming highlighted and non-highlighted Other pictures in the deceptive and switch Context. Error bars represents 95% bootstrapped confidence intervals. Predicted mean proportions from the mixed-effects logit model (stars) and horizontally-jittered by-subjects observed proportions are also displayed.

**Naming latencies**

Naming latencies for the 32 critical pictures were collected. We excluded naming errors (1.1%), disfluencies (3.7%), and trials when the recording device failed detecting latencies (8.5%). Latencies shorter than 250ms (0.2%) or longer than 2500 (0.7%) were also excluded. The analysis was based on 1319 data points (Truthful: 415; Deceptive: 460; Switch: 444).

Latencies were log-transformed for analysis as this addressed both heterogeneity and non-normality of residuals. Figure 13 shows density plots of latencies per condition before and after transformation. Table 6 reports the descriptive statistics for raw RTs split by Context.
and Case, and Target Response respectively.

We built two linear mixed-effects models. The first model included Context (treatment-contrast coded with Truthful as baseline) as only predictor. The second model was restricted to data in the deceptive and switch conditions, and included Context (Deceptive = .5; Switch = -.5), Case (Other = .5; Highlighted = -.5) and their interaction as fixed effects. As the dependent variable was log-transformed, the exponentiated regression coefficients represent the ratio of the expected geometric means (ratio-gM) of RTs between the two levels of a predictor. The significance of the fixed-effects parameters was assessed via likelihood ratio test (LRT) comparing the models with and without the parameter of interest, following Levy (2014).

Relative to participants in the Truthful condition, participants in both the Deceptive (β = 0.18, SE = 0.06, t = 2.89, ratio-gM = 1.20, 95% profile CI [1.06; 1.36], LRT p = .002, 1 d.f.) and Switch (β = 0.16, SE = 0.06, t = 2.53, ratio-gM = 1.17, 95% profile CI [1.04; 1.33], LRT p = .011, 1 d.f.) group took significantly longer to name pictures.

---

18 The model included by-subjects random intercept, and by-items random intercept and slopes for Context. Model df: 1319 N, 48 subjects, 32 items. All correlations between fixed effects < |.7|
19 By-subjects random intercept and slope for Case, and by-items random intercept and slopes for all fixed-effects parameters were included. Model df: 904 N, 32 subjects, 32 items. All correlations between fixed effects < |.15|.
We found no difference in naming latencies between participants in the switch and the deceptive condition ($\beta = 0.02$, SE = 0.06, $t = 0.38$, ratio-$g_M = 1.02$, 95% profile CI [0.91; 1.15], LRT $p = .693$, 1 d.f.). Naming the (non-highlighted) other vs. the highlighted picture did not affect RTs ($\beta = -0.05$, SE = 0.03, $t = -1.58$, ratio-$g_M = 0.95$, 95% profile CI [0.89; 1.01], LRT $p = .112$, 1 d.f.), regardless of whether participants were naming the non-highlighted picture in an attempt to deceive or benefit their partner ($\beta = 0.006$, SE = 0.06, $t = 0.10$, ratio-$g_M = 1.01$, 95% profile CI [0.89; 1.13], LRT $p = .913$, 1 d.f.).

Table 6

<table>
<thead>
<tr>
<th>Case</th>
<th>Truthful</th>
<th>Switch</th>
<th>Deceptive</th>
<th>(tot)</th>
</tr>
</thead>
<tbody>
<tr>
<td>other</td>
<td>952 (297)</td>
<td>986 (326)</td>
<td>969 (312)</td>
<td></td>
</tr>
<tr>
<td>highlighted</td>
<td>1019 (361)</td>
<td>1038 (370)</td>
<td>1029 (365)</td>
<td></td>
</tr>
<tr>
<td>(tot)</td>
<td>843 (287)</td>
<td>986 (332)</td>
<td>1012 (349)</td>
<td></td>
</tr>
</tbody>
</table>

Cognitive Load (memory test)

Participants' performance on each of the 9 memory tests were coded as correct if all the four previously-unselected objects were identified, incorrect otherwise. This resulted in a dataset of 432 data points. Accuracy (correct = 1, incorrect = 0) was our binomial outcome variable.

We built a mixed logit model with Context (treatment-contrast coded with Deceptive as baseline) as fixed-effects predictor, and by-subjects random intercept and by-blocks random intercept and slopes for the two contrast comparisons. Truthful and Switch groups were compared using glht. We found no difference in performance across the three Context conditions (see Table 7): Truthful vs. Deceptive ($\beta = 0.21$, SE = 0.42, $p = .629$, odds-ratio $= 1.23$, 95% profile CI [0.52; 2.94]), Switch vs. Deceptive ($\beta = -0.39$, SE = 0.44, $p = .366$,...
odds-ratio = 0.67, 95% profile CI [0.27; 1.62]), Truthful vs. Switch (β = 0.60, SE = 0.43, p = .167, odds-ratio = 1.82, 95% CI [0.78; 4.25]).

Table 7

Frequency and proportion ±SE of Correct (incorrect) responses on the memory tests per Context conditions

<table>
<thead>
<tr>
<th></th>
<th>Truthful</th>
<th>Switch</th>
<th>Deceptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>49 (95)</td>
<td>35 (109)</td>
<td>44 (100)</td>
</tr>
<tr>
<td>% Correct</td>
<td>34% ±12%</td>
<td>24% ±11%</td>
<td>31% ±12%</td>
</tr>
</tbody>
</table>

Feeling of Closeness

Participants' responses on the IOS scale are displayed in Figure 14.

![Figure 14](image)

*Figure 14* Frequency of responses on the IOS scale (least close = 1; most close = 7) split by Context.

Feeling Of Closeness was the ordinal dependent variable used in this analysis. We used the `polr` function from the `{MASS}` package in R (Venables & Ripley, 2002) to estimate an
ordered logistic regression with Context (treatment-contrast coded with Deceptive as baseline) as predictor. Exponentiated coefficients represent proportional odds-ratio. We controlled for Gender and Confederate. To cope with the issue of sparse cells (i.e., cells with no observations which may bias the model estimates), we removed level 7.

In relation to participants in both the Switch and Truthful Context, deceivers felt less close to their partner: Truthful vs. Deceptive ($\beta = 1.45$, SE = 0.69, $p = .034$, prop. odds-ratio = 4.28, 95% profile CI [1.15; 17.21]), Switch vs. Deceptive ($\beta = 1.97$, SE = 0.71, $p = .005$, prop. odds-ratio = 7.20, 95% profile CI [1.88; 30.42]). Neither Gender ($p > .8$) nor Confederate ($p > .6$) was significant.

However, when Feeling of Closeness was included in the mixed logit model for convergence data together with its interaction with Context, it did not predict the likelihood to lexically converge in either Context condition (all $p$s $> .2$).

### 3.8.3 Discussion

In contrast to our expectations, we found no significant differences in participants’ likelihood to lexical converge across the three Context conditions. Importantly, we did not replicate the patterns observed in Experiment 2: deceivers did not converge less than truth-tellers. Furthermore, convergence tendencies were the same when participants named the highlighted vs. the non-highlighted pictures, regardless of whether they were naming the non-highlighted picture to deceive or benefit their partner. Participants were significantly slower at naming pictures in the Deceptive and Switch Context compared to the Truthful Context condition, a pattern that suggests a task-switching cost. However, in both the Deceptive and Switch conditions, participants were as fast at naming non-highlighted as highlighted pictures, thus casting doubts about the task-switching cost explanation. In contrast with what was observed in Experiment 2, deceiving did not hamper participants’ performance on the memory test. Similarly, there was no effect of simple task-switching (Switch vs. Truthful comparison). Finally, participants in the Deceptive condition reported greater psychological detachment toward their interaction partner than participants in either Truthful condition, in line with what the social-less hypothesis would predict. However, how close they felt did not predict the extent to which they lexically converged.
3.9 **GENERAL DISCUSSION**

In three Experiments, we investigated whether having deceptive intents towards an interlocutor, and lying to them, could interfere with the processes of linguistic convergence found so far between truthful dialogue partners. We focused on lexical convergence and used two interactive priming paradigms in which participants were primed by their partner with appropriate but dispreferred names to refer to objects. Half of the participants were asked to deceive their partner. We operationalised deception as the intention (and lying as the subsequent act) to instil a false belief in the other's mind. Similar to many real-life scenarios, our deceptive manipulation unfolded as a social exchange with a naïve other (in Experiment 2 and 3 the confederate was naïve about the main manipulation and purpose of the study). Importantly, from the point of view of the deceiving participants, their deception had clear repercussions on the interaction and the actions/performance of the deceived (i.e., the Investigator identified the wrong culprit, the other performed poorly on the memory test).

We did not find any consistent evidence that deception or lying affected the degree of lexical convergence between deceivers and deceived, at least not to distinguish it statistically from the degree of lexical convergence between truthful interlocutors. As we discussed at the end of Experiment 1, the detective-scene paradigm used therein may have represented a too hard task for the deceptive and honest Respondents who had to figure out in a matter of seconds who was the culprit. With most of their resources engaged in this task, participants may have fewer left to monitor the appropriateness of the words used by their partner, leading to convergence effect almost at ceiling level and preventing any modulatory role for any of our manipulation in either condition. We also saw that, once in the context of a scene, the original preferred/dispreferred name categorisation might have no longer been in place. Participants may have converged to such an extent because the primed name would have been their personal choice for that object.

We addressed both issues in Experiment 2 and 3, using a simpler picture-naming/-matching paradigm with decontextualized pictures of objects. In Experiment 2, we observed weaker...
lexical convergence tendency in deceivers compared to truth-tellers. Deceivers also showed greater effects of cognitive load by performing worse on the memory test. However, none of these findings was replicated in Experiment 3 (where we investigated the nature of the cognitive cost that presumably decreased deceivers’ convergence in Experiment 2).

It could be argued that the paradigm employed in Experiment 2 and 3, with participants instructed when to lie, prevented them from really engaging in deception (i.e., switching to a deceptive mode). However, note that even if the instruction mode might have spared deceivers from the decision-process to lie, they still had to suppress a prominent true stimulus and produce a conflicting false response: processes that independently add up to the demands of lying (see section 3.3.1). In addition, as already said, participants were aware of the consequences of their action on their unaware partner, and so we should expect emotional load to be present. For these reasons, we argue against this possibility.

Thus, can we conclude that there is no genuine effect of deception on lexical convergence? If we look at the estimated effect sizes and related confidence intervals for the deception manipulation in the three experiments (odds-ratio: [0.41; 1.76], [1.01; 1.45], [0.53; 2.28]), we see that they are rather narrow and have extremes that are not far from one (i.e., all effect sizes in the intervals are small). This suggests that the effect, if there, is small. In Experiment 3, where we failed to replicate the effect of deception on lexical convergence found in Experiment 2, participants in the two task-switching conditions (Deceptive and Switch) showed corresponding signs of cognitive load (their naming latencies were significantly slower than those of truth-tellers). In addition, deceivers also reported feeling psychologically less close to their dialogue partner compared to participants in either of the two truthful conditions. These two results suggest that some of the processes that could potentially interfere with lexical convergence mechanisms according to both families of accounts (i.e., impaired attentional resources, disaffiliation) were in place. However, they were not strong enough to actually interfere with the convergence mechanisms. Why could this be?

One possibility, linked to the social accounts, is that deceiving may not differently affect the “strategic” use of lexical choices than truth-telling in highly cooperative settings. We can think of two reasons. First, our truthful context was not neutral: the participant was actively
contributing to the success of the pair's joint (Experiment 1) or the other's (Experiment 2 and 3) task. It may be that this highly cooperative setting pushed truth-tellers to accommodate to the unusual lexical choices of their partner to a greater extent than they would had done under truthful yet less cooperative circumstances. Truth-tellers thus ended up converging to a similar extent as deceivers even if they did so as a result of very different social processes (deceivers did so to ingratiate with their victim). Adding a more neutral truthful condition could clarify whether both truth-tellers’ and deceivers’ lexical convergence rate was socially affected in the current task. The second reason is linked to the fact that a speaker's lexical choices are more explicitly controlled and more effortful than other aspects of their speech (e.g., accent or grammar) (Greenwald & Banaji, 1995). In line with this claim, Scissors, Gill, and Gergle (2008) found that participants who defected more often in an interactive economic game, and so betrayed their partner's trust for personal benefits, showed a higher degree of linguistic convergence on lower-level aspects of language production (e.g., standard responses, like 'yeah') but not on content-related words (e.g., nouns). Deceivers, like defectors, are probably associated with depleted cognitive resources as both behaviours require a certain amount of cognitive effort. Thus, our deceivers, like Scissors et al.’s defectors, may have “fewer resources to attune to a partner's language and thus employ less effortful forms of mimicry rather than more effortful, content-related mimicry” (Scissors, et al., 2008, p. 280).

A second possibility is linked to our between-subjects manipulation and the fact that in our study, as in most studies on linguistic convergence, we did not measure the individuals' baseline preference for using either alternative name used to operationalise lexical convergence. Looking forward, measuring this preference in advance, during a separate pre-experimental session, would allow for a refined operationalisation of lexical convergence as the change in the individual probability to produce either name due to the experimental manipulation. Alternatively, experiments should probably move away from what is likely to be perceived as an artificial choice between names by participants (see our attempt in Chapter 5).

Related to our previous points is the question of the individual differences that may affect people's natural propensity to linguistically converge with others. Compatibly with Pickering
and Garrod's (2004) IAM, individual differences in attention and executive control should modulate participants' alignment tendencies. Following Giles et al.' (1991) CAT, social skills like propensity to compromise and dominance during conflict (as suggested in Weatherholtz et al., 2014) should equally modulate linguistic convergence. Future studies, especially those relying on between-subjects designs, should account for these potential effects and run experiments with appropriate sample sizes.

Finally, audience-design dynamics (cf. Fussell & Krauss, 1992) may have played a role in our experiment given our operationalisation of lexical convergence as the choice between alternative names (one of which was highly dispreferred). On the one hand, the unusual choice of name may have led participants, independently of the condition, to think in terms of communication efficiency. This led them to opt for the name that they thought the other speaker would more easily and quickly respond to (i.e., the one they had used), cancelling out any possible effect due to our experimental manipulation. Alternatively, and more so in our picture-naming paradigm, participants may have realised that it did not really matter how they named the objects: the object to be named was clearly different from the distractor and even not-so-appropriate terms would have enabled the selection of the right picture. Thus, the name choice may not have been perceived as essential to the pair's communicative goals, an aspect which is important for linguistic convergence to occur according to those audience-design explanations for which convergence is the result of a conceptual pact (Clark, 1992; Carbary & Tanenhaus, 2011).

However, considerations of sample size and experimental paradigms might not fully account for our results. We should in fact acknowledge that, across the three studies, speakers showed lexical convergence with their interlocutor's choice, independently of whether they were deceiving their interlocutor by sometimes lying to them or were instead always telling them the truth. Across the three experiments, the proportion of times speakers used the dispreferred names used by their partner (between 53% and 76%) was well above the proportion of speakers who chose the dispreferred name in our forced-choice stimuli pre-test task (i.e., below 20% of times). This provides evidence against a straightforward social-modulated explanation of lexical convergence. Instead, our results suggest that lexical convergence may be robust to modulation in the social context; at the same time, they leave
open the possibility that lexical convergence may be mediated by non-social mechanisms, in line with the theories that see convergence as an automatic process (e.g., Pickering & Garrod, 2004).

3.10 CONCLUSIONS

Data from this study are inconclusive. In three experiments, using two novel interactive priming paradigms, we did not find any consistent evidence that deception or lying affected the degree of lexical convergence between deceivers and deceived, at least not to distinguish it statistically from the degree of convergence between truthful interlocutors.
CHAPTER OVERVIEW FOR CHAPTERS 4 AND 5

The aim of the research reported in the following two chapters was two-fold. At a theoretical level, we focused on the social accounts of linguistic convergence (e.g., Giles & Coupland, 1991) and investigate two aspects that have been suggested to affect the degree of linguistic convergence between interlocutors (see Chapter 2): the use of convergence as a tool to manage social distance; the consequences on convergence of contextual social variables such as the impressions that speakers form of their interlocutor and the motivational goals of the interaction (i.e., goal-oriented vs. less-goal-oriented dialogues). We tested whether speakers were more likely to converge with the lexical choices of a partner who, on a previous occasion, had shown trusting and collectivist behaviour towards them compared to a partner who had shown mistrust and an individualistic attitude. We checked this possible effect of partner perception on lexical convergence in two sub-tasks characterized by different degrees of joint goal, to see whether the motivation of a joint goal further modulated convergence.

But our motivation was also (and foremost) methodological: we wanted to explore the potential use of cross-recurrence quantification analysis (CRQA) for the investigation of lexical convergence in laboratory-based conversation. Unlike the artificially controlled lexical convergence lab tasks used in Chapter 3 and in many dialogue experiments in psycholinguistics before (see Chapter 2), participants in our Chapter 5 experiment engaged in an unconstrained dialogue task. As in real-world conversations, they were free to structure their linguistic exchange as they pleased, and were not forced to simply choose between two alternative object names. As a result, we could not measure lexical convergence as the likelihood to re-use the previously heard alternative. The conversation between our participants unfolded in unforeseeable yet not fully random manners, and showed characteristics of non-stationarity and non-linearity typical of spontaneous “real” dialogues and, more generally, complex systems. In Chapter 4, we thus present CRQA as a possible solution to this quantification challenge and show how it can be successfully used to measure
not just the magnitude, but also the stability, the variability and the temporal evolution of linguistic convergence in conversations. A large part of this chapter is thus dedicated to the discussion of the possibilities that this technique opens for laboratory-based dialogue research; at the same time, we attempt to provide an overview of the issues and decisions the researcher must face at implementation and offer some solutions.

Hence, these chapters are structured as follows: first, in Chapter 4, we discuss CRQA and the methodological factors that inspired our study. Second, in Chapter 5, we outline the framework that motivated it theoretically, and present the experimental method and findings.
Chapter 4

CROSS-RECURRENCE QUANTIFICATION ANALYSIS: A FRIENDLY BUT CRITICAL INTRODUCTION

4.1 CROSS-RECURRENCE PLOTS (CRPs) AND CROSS-RECURRENCE QUANTIFICATION ANALYSIS (CRQA) FOR DISCRETE TIME SERIES

In this section, we present the basic concepts of categorical Cross-Recurrence Plots (CRPs) and Cross-Recurrence Quantification Analysis (CRQA), the statistical tools that we will use for the analysis of our linguistic data. We only present the approach for discrete time series as it is the approach applied in our study (for a comprehensive introduction to the method, see Marwan, Carmen Romano, Thiel, & Kurths, 2007; Webber & Zbilut, 2005; for its applications to continuous time series see Marwan & Webber, 2015; Schmidt & Richardson, 2008).

4.1.1 (Verbal) interpersonal interaction as bivariate discrete time series

Interpersonal interaction, like many other aspects of human behaviour, is often studied by recording multiple observations of the same behaviour (statistically, a random variable) at regular intervals of time. Hence, interaction research usually deals with time series of behavioural data. Technically speaking, a time series is the observed “value” of a random process \( X(t) \) such that at any time \( t_i \), \( X(t_i) \) is a random variable. A random process is discrete-time if \( t \) comes from a countable set of time points \( t \in \{0, 1, 2, \ldots, n\} \). Importantly, even if the sampling of \( X \) is ordered in time, the information about the exact onset and duration of these times is not used (for this reason, discrete-time random processes are also called “random sequences”). \( X \) can then be continuous (e.g., the pitch of a speaker's voice – as fundamental frequency – at the onset of utterances) or discrete (e.g., the number of prepositional-object structures in a person's speech). The former will be a case of
continuous-valued discrete-time random process, the latter of discrete-valued discrete-time random process.

Sampling the same behaviour regularly over time allows researchers to study how this behaviour (e.g., the use of prepositional-object structures) changes over time (e.g., over the course of a conversation), perhaps as a function of other factors (e.g., topic or interlocutor). For behaviours like interpersonal interaction and language use in conversation, the time dimension is an intrinsic component of the behaviour itself, i.e., all behaviours necessarily involve, and can be repeatedly recorded over time. However, while we could gain a meaningful measurement and understanding of, for instance, an individual’s level of arousal or muscle activity at a single point in time, such a single-time measure would be impossible to obtain and interpret for a behaviour like the linguistic exchange between two individuals which requires historical information: what has been said until then. Understanding these behaviours thus requires investigating how they unfold over time.

In this chapter, we will be dealing with bivariate discrete-valued discrete-time random processes: the lexical production of two dialogue partners. Trivially, when a person speaks, they produce a sequence of words: a discrete time series of multinomial observations (multinomial as choice of words is discrete but not binary). Similarly, we can see the conversation between two people as two such random processes interacting over time. We can ask many questions about these time series; for instance, whether one is more likely to contain a prepositional-object structure following the occurrence of a prepositional-object structure in the other. Importantly, we need statistical methods that are able to capture how these separate streams are associated as they each evolve over time.

4.1.2 Why is CRQA suitable for the analysis of conversations?

A speaker’s verbal production and, more broadly, verbal (and non-verbal) interactions tend to be noisy, and behave in non-linear and non-stationary manners (Coco & Dale, 2014; Dale,

22 Different aspects of a person's speech production can of course be represented by different types of random processes. For instance, if we measure the pitch of the speaker's voice (a continuous variable) on the first phoneme of each utterance, we will be dealing with a continuous-valued (i.e., fundamental frequency) discrete-time random process. If we continuously record the speaker's pitch throughout a whole utterance, we will be dealing with a continuous-valued continuous-time random process.
Warlaumont, & Richardson, 2011). In simple terms, a time series is linear and stationary (in a strong sense) if its statistical properties (i.e., the expected mean, variance, covariance and autocorrelation of the behaviour being measured), do not depend on the time at which the series is measured (Chatfield, 2004). Going back to our definition of a random process, stationarity means that for any time \( t_i \) and \( n \), \( X(t_i) \) and \( X(t_i+n) \) share the same probability distribution. White noise is a classic example; it is not relevant when you measure it, it should look pretty much the same at any period of time. In contrasts, time series with trends or seasonality, are not stationary: trend and seasonality will change the time series’ statistical values at different times (e.g., in upward-trended time-series the mean increases over time). Similarly, two time series are jointly stationary if each is stationary and their cross-correlation (i.e., a measure of the dependency between their observations) only depends on the chosen lag but not on the sampling time.

It follows that a conversation would be linear and stationary if it evolved (or, better, non-evolved) in time like a minimalist music piece: an unending repetition\(^2\). But a conversation is not such a collection of endlessly predictable wording. Even if studies have shown that speakers can predict many aspects of their interlocutor's upcoming speech (e.g., Pickering & Garrod, 2013) and adjacent utterances tend to be topic-coherent (Grosz & Sidner, 1986), over the long run, unconstrained dialogues evolve in unforeseeable manners and word choices cannot be predicted from the start. Conversations are neither fully random nor deterministic: what a speaker says and how she says it are deeply coupled with what both she and her interlocutor have said, in a way that guarantees speakers are responsive to each other's contributions and builds towards share understanding; at the same time, however, their linguistic production is flexible enough to take unexpected turns as a result of both internal (e.g., you got upset by a sudden thought) and external (e.g., your flight departure is announced and you need to rush to the gate) factors. Even so, a conversation is never a fully balanced mutual exchange (Gorman, Cooke, Amazeen, & Fouse, 2011); speakers alternate as speaker and listener, and how smoothly they coordinate their turn and how long they each speak, with their own personal linguistic style, are all factors affecting the time evolution of their linguistic production.

\(^2\) No judgment intended, I am a fan of the genre.
The fact that human interaction is non-linear and non-stationary does pose methodological challenges to those who want to study it in quantitative ways (Fusaroli, Konvalinka, & Wallot, 2014). In fact, unless its statistical properties like means and cross-correlation are fixed or predictable rather than evolving in an unknown manner, it won’t be possible to sensibly average the observed data. For instance, cross-correlation has been shown to be biased (i.e., producing larger coefficients) if the two time series are not stationary or have not been preventively ‘whitening’ (i.e., extract the stationary time series driving it; Chatfield, 2004). More importantly, non-stationarity is a characteristic of interpersonal interaction. As such, accounting for it (instead of trying to clean it away) could further our understanding of interpersonal interaction itself.

Cross-Recurrence Quantification Analysis is a statistical technique borrowed from dynamical-systems where it has long been applied to investigate the inter-relations of complex systems (Zbilut, Giuliani, & Webber, 1998; see Marwan, 2008 for an historical overview). CRQA works by reconstructing the phase space of two systems simultaneously (i.e., a graphical representation of all the possible states of the systems, with each possible state corresponding to one unique point in the space) and tracing the systems' actual trajectories over time (i.e., the actual states visited by each system). All the occurrences of when the two systems re-visited the same or a very similar state over time are then recorded in a matrix and cross-recurrence plot (CRP; Marwan & Webber, 2015). CRQA measures the extent and the characteristics of these patterns of co-recurrence in the CRP.

It is easy to see how such a statistical tool is appealing to the study of interpersonal interaction and conversation; due to its capacity to objectively quantify the extent and form of the patterns of co-visitation between two systems, it can measure the degree with which two interlocutors' speech production exhibits similar patterns (e.g., converges on the same words), and how complex and extended (vs. regular and limited in time) their pattern of convergence is (see Shockley & Riley, 2015 for a review).
4.2 CATEGORICAL CRQA FOR THE STUDY OF LEXICAL CONVERGENCE IN CONVERSATION: HOW DOES IT WORK?

In our study, we used CRQA to characterise and quantify the patterns of lexical convergence in participants’ conversations in a ‘plane crash’ experiment, to be fully described in Chapter 5. In this section, we outline how we can apply CRQA to this type of data and what information about lexical convergence we can extract from it. We start by outlining how categorical CRPs are constructed and then explain the CRQA measures that describe them.

Each interlocutor's lexical production is turned into a sequence of mutuallyexclusive numerical codes, each code representing one unique word. Codes are ordered in time mirroring the order with which the corresponding words were produced. Yet the sequence contains no information about the actual onset and duration of the produced words (recall that a two-person conversation can be thought as a bivariate discrete-time multinomial time-series). If the conversation has a strict turn-taking structure (i.e., speakers’ talk never overlaps), we recommend modelling the temporal structure of turn-taking by coding speakers' silent turns (i.e., inserting dummy code in speaker A's sequence when speaker A is silent and speaker B is talking). Failing to model silence would in fact artificially inflate cross-recurrence (see section 4.2.1).

Here is a simple illustrative example; a fictitious dialogue between Lucy and Linus.

(Dialogue 1A)

Lucy: Gosh, you keep repeating people.
Linus: Keep repeating people? Sorry, what?
Lucy: See! You keep saying what I say!
Linus: I don't keep saying what I say.

By assigning a number to each unique word (e.g., “keep” = 6), the word sequences are turned into discrete numeric sequences. Note that the magnitude of the numbers bear no meaning; the data is nominal.

(Dialogue 1B)

Lucy: 2 12 4 6 5
Linus: 4 6 5 10 11
Lucy: 9 12 4 8 11 3 7

131
Adding dummy code for silent turns (i.e., 888 and 999) yields two numeric sequences, each representing the speech stream of one speaker in the conversation.

(Dialogue 1C)

Lucy: 2 12 4 6 5 888 888 888 888 9 12 4 8 11 3 7 888 888 888 888 888 888 888 888 888 888 888 888 888

Linus: 999 999 999 999 999 4 6 5 10 11 999 999 999 999 999 999 999 3 1 4 8 11 3 7

The discrete numeric sequences of Dialogue 1C are the input for a CRP\textsuperscript{24}. A CRP is built as follows. If the two sequences are of same length N (as it is usually with conversations in which silent turns were dummy-coded, and, more generally, whenever participants' behaviour is sampled at the same points in time for an equal number of times)\textsuperscript{25}, each speaker's word sequence is plotted on one of the axes of a N×N recurrence plot. The main diagonal line (line of synchronization) represents behaviours that occur exactly at the same time in the two time series. It follows that it is excluded from the estimation of cross-recurrence of turn-taking behaviour.

A pixel is filled in the plot at those coordinates in which speaker-X(j) and speaker-Y(i) have the same value – in our example, when Linus produces a word that was previously produced by Lucy (and vice versa). Technically, this is achieved by fixing the threshold of recurrence (radius) at 0 and the embedding dimension at 1, so that a discrete code (word) in a speaker's sequence must exactly match the discrete code (word) in the other speaker's sequence\textsuperscript{26}. The

\textsuperscript{24} Categorical CRQA is a “lucky case” compared to continuous CRQA because it does not require the re-construction of the phase space of the underlying systems. This is achieved by choosing an appropriate value for the embedding and the delay parameters. In the case of discrete-valued discrete-time time series, previous studies have shown that embedding = 1 and delay = 1 are accurate dimensions (Coco & Dale 2014; Dale et al. 2011).

\textsuperscript{25} If the two sequences are of different length N and M, with M > N, then a N×M matrix is constructed (this is usually the case with continuous-value time series sampled with different scales, but in this case the line of sync must be approximated). Alternatively, if the difference is within a certain small threshold (e.g., 10 units), then the longer sequence is cut off to match the length of the shorter sequence (and a squared N×N matrix can be formed). This latter is the option suggested in the \{crqa\} R package for categorical time series.

\textsuperscript{26} For continuous-valued time series, two points in the time series are considered as (cross-)recurring if they fall within a certain radius (with an upper threshold not too far from zero). Due to the nature of continuous data in fact, recurrence cannot be estimated via exact mismatch/match between the
filled-in pixels are referred to as recurrence dots. For our purposes, a CRP is thus a structure that can capture where, over the course of a conversation, the same words are being repeated between the two interlocutors. We can then apply CRQA to quantify these patterns of repeated word in the CRP (see Figure 15 for more details). Importantly, this quantification does not tap into the meaning of the repeated-word patterns, but more superficially captures the “way of talking” of two interlocutors (Franco Orsucci et al., 2006) (e.g., in the following exchange between Steve and Jamie, the repetition of “break out” would count as two recurrence dots, despite the very different meanings implied. Steve: “You can tell what’s causing your acne by where you break out. Mine is due to stress by over-control.” Jamie: “Let’s break out the champagne then!”). The “cruder” measure of convergence between two speakers' lexical production is cross-recurrence rate (RR): the proportion of recurrence dots in the CRP as percentage of all possible dots. It captures the extent to which the speakers are using the same words. However, it also includes repetitions between speakers that may not be the result of linguistic convergence mechanisms (i.e., “chance recurrence” or “spurious convergence”); for instance, the isolated dot corresponding to the repetition of “what” between Linus and Lucy at c(15,10) in Figure 15. Note though that isolated dots are not always a sign of noise (Marwan & Webber, 2015).

More accurate for the investigation of lexical convergence are the measures that characterize the patterns of recurrence dots that form diagonal lines parallel to the line of sync. These represent in fact repetitions of at least two words in a row and so capture those cases in which speakers' lexical productions do not merely match each other accidentally, but exhibit continued entrainment with each other (e.g., repeating phrases or even whole sentences; cf. Gorman et al., 2011).

The closer a diagonal line is to the line of sync, the closer in time a speaker repeated what the other had said (Dale & Spivey, 2006); thus, they can be interpreted as cases of “local” or “short-term” linguistic convergence (i.e., a speaker producing sequences of words that matched those employed by their interlocutor in a closely preceding utterance). Diagonal
lines that are far from the line of sync represent repetitions that happened far away in time during the conversation, and may thus be considered more “global” forms of linguistic adaptation (i.e., long-term adaptation or priming effects, usually indicative that interlocutors have converged on a common language and discourse representation; e.g., speakers’ developed a common reference system). Overall, the more entrained two speakers are, in the sense of sharing the same language use, the more recurrence dots will be organised along diagonals (cf. Fusaroli et al., 2014). Determinism (DET) captures the percentage of recurrence dots that lie on diagonal lines and thus constitutes a more accurate measure of convergence than RR. Furthermore, unlike RR, DET is not systematically affected by modelling silent turns as explain below (see also Figure 16/Figure 17).

The average diagonal length (L) represents the average duration (in word units here) that speakers tend to repeat each other. Highly predictable and routinised conversations, in which speakers tend to quickly converge on a common standardised way of speaking and/or rely on linguistic routines, will have long diagonal lines, and so long L, as their speech show prolonged sequences of shared words. For instance, when people first meet, their initial exchange tends to follow precise patterns (e.g., imagine Steve and Jamie are introducing each other at a party for the first time. Steve: “Hi, I’m Steve. Nice to meet you.” Jamie: “I’m Jamie. Nice to meet you, Steve.” Steve: “What do you do for a living, Jamie?” Jamie: “Oh, I’m doing a PhD. I’m not exactly sure it is ‘for a living’ though! What do you do?”). Vice versa, short diagonals and/or individual dots characterise more unpredictable or fluctuant conversations (cf. Webber & Zbilut, 2005). Related to L is the length of the longest diagonal line (maxL) excluding the main diagonal line. Importantly, the shorter L and maxL, the less stable the conversation is in terms of repeated language use (Zbilut et al., 1998).

Finally, the Shannon entropy of the frequency distribution of the diagonal line lengths (ENTR) is related to the complexity of the repeated linguistic patterns. For instance, interlocutors that converge on different types of linguistic structures (e.g., from chunks of two words, to longer phrases and sentences of various length) will show high ENTR; in contrast, interlocutors that only converge on the same few types of linguistic structures will show low ENTR, characterising their very regular attunement (cf. Fusaroli et al., 2014). If all the diagonal lines are of the same length, then ENTR is zero. Relative entropy (rENTR) is the
entropy measure normalized by the number of lines in the CRP (for instance, if all the diagonals in the plot are of different length, $r_{ENTR}$ will be 1). As such, *Relative entropy* is preferred when comparing entropy across contexts and conditions (Coco & Dale, 2014).

Finally, it is important to emphasise that all diagonal-based measures, being based on $\%RR$, are normalised with respect to the length of the conversation and number of words produced.

![Figure 15 Illustration of the cross-recurrence plot (CRP) for Dialogue 1A at word level. The main diagonal line represents simultaneous time (i.e., $t_{Lucy} = t_{Linus}$); as our speakers could not talk simultaneously, it was removed from the estimation of cross-recurrence upon plot construction (Theiler window parameter = 1). Leader-follower dynamics can be captured in the upper and lower triangular areas. In the upper triangular area, Linus’ (i.e., the speaker on the y-axis) word production always follows Lucy’s (i.e., the speaker on the x-axis). In fact, $t_{Linus} > t_{Lucy}$, for any $t$. Hence, recurrence dots represent Linus converging on Lucy’s lexical choices. In the lower triangular area, Lucy’s (i.e., speaker on x-axis) word production always follows Linus’s (i.e., speaker on the y-axis), as $t_{Lucy} > t_{Linus}$, for any $t$. Hence, recurrence dots in this lower area represent the extent to which Lucy is converging on Linus’ word production. In this example, we can see that Linus converged with Lucy’s word choices to a bigger extent than the other way around; the upper triangular area contains in fact many more recurrence dots (i.e., $RR_{upper} > RR_{lower}$). CRQA measures: $\%RR = 2.08$, $\%DET = 66.67$, $L = 4$, and $maxL = 5$. $ENTR$ is 0.69 and $r_{ENTR}$ is 1 (as the two diagonals on the plot are of different length).
4.2.1 Exploring categorical CRPs and lexical coding a bit further

Let's see what happens if we do not dummy code silent turns, thus violating the time constraints on the conversation due to turn-taking. We do this when excluding the main diagonal from the CRP (Figure 16) and when including it (Figure 17).

Here are the numeric sequences used for the construction of the CRP of Dialogue 1A if we do not dummy-code silence:

(Dialogue 2C)  
Lucy: 2 1 2 4 6 5 9 12 4 8 11 3 7  
Linus: 4 6 5 10 11 3 1 4 8 11 3 7

*Figure 16 CRP plot of Dialogue 1A without dummy-coding for silence, removing line of synchronization from construction*

The CRP plot is not correctly capturing who is converging with whom (Lucy results converging with words that were produced by Linus after she did). This is reflected in the 'spurious' diagonal line “what I” (10,11):(5,6) which also inflates %DET.

As we excluded the main diagonal line (to accurately model the fact that speakers never talked simultaneously), we failed to capture the full sentence repeated verbatim between speakers (keep saying what I say).

CRQA measures: %RR = 4.86, %DET = 71.42, L = 2.5, maxL = 3, ENTR = 0.69, rENTR = 1

*Figure 17 CRP plot of Dialogue 1A without dummy-coding for silence, including line of synchronization from construction*

The CRP contains a “fake” line of sync as the speakers are actually never talking simultaneously.

Here as well the CRP is not correctly capturing who is aligning with whom.

This time we are able to capture the alignment on the longer sentence, but this is misleadingly represented on the line of synchronisation.

CRQA measures: %RR = 8.33, %DET = 83.3, L = 3.33, maxL = 5, ENTR = 1.10, rENTR = 1
Figure 16 and Figure 17 show how modelling the actual temporal dynamics of the conversation by representing speakers’ silent turns via dummy-coding is important to correctly capture who is lexically converging with whom, thus preventing false recurrence dots and diagonals (which inflate $RR$ and $DET$). However, this also means ending up with rather sparse recurrence matrices, by constructions. $RR$ is in fact higher when we do not dummy-code silence as there are fewer possible dots overall (the matrix is smaller). This is especially so along the bands close to the main diagonal. As a confirmation,

Figure 18 shows the CRP plot for the following extreme-case conversation, in which Linus repeats verbatim Lucy. $RR$ is only 5%. However, $DET$ is 100%, which correctly reflects the degree of convergence between the speakers.

Lucy: Stop copying everything I say!
Linus: Stop copying everything I say!
Lucy: Stop copying everything I say!
Linus: Stop copying everything I say!

Figure 18 %RR = 5; %DET = 100; L = 5; $maxL = 5$; $ENTR = 0$; $rENTR = NaN$
4.2.2 Windowed (categorical) CRQA

The CRQA presented in the previous section allows quantifying various characteristics of two speakers' linguistic convergence over the course of their interaction. However, the estimated measures characterise their overall level (i.e., magnitude, stability and regularity) of convergence. The analysis per se does not capture how this convergence may have evolved as the interaction unfolds. Windowed CRQA captures exactly that: by estimating CRQA measures in overlapping windows (of a specified size) sliding along the main diagonal, it tracks how lexical convergence changes as the interaction progresses. For each window, a CRP plot is constructed and CRQA measures are estimated (Coco & Dale, 2014). Adjacent windows are sampled at a specified interval so that they may overlap to greater or lesser extents. Importantly, because the windows are smaller than the whole conversation and slide along the line of synchronization, windowed CRQA manages to capture matching lexical choices that happen in relative temporal proximity rather than across the entire length of the conversation and may thus more accurately measure short-term linguistic convergence.

As can be seen in the example (Figure 19), results very much depend on the choice of window size and sampling interval. In particular, too small window sizes may prevent us from estimating certain CRQA measures (e.g., DET and ENTR) because they do not capture enough data. On the other hand, too big window sizes may be more likely to include “spurious” cross-recurrence, or cases of distant repetitions that most likely capture more long-term lexical adaptations. In section 5.2.7.1, we propose a method for selecting suitable values for window size and step based on maximizing the signal-to-noise ratio.

Note that windowed CRQA is different from the diagonal-wise cross-recurrence profile approach used in some studies (e.g., Richardson & Dale, 2005). Windowed CRQA estimates all CRQA measures for each window. Diagonal-wise profiling only estimates RR.
4.2.3 Categorical Recurrence-Quantification Analysis (RQA) applied to conversation

Some authors (e.g., Fusaroli & Tylén, 2016) have proposed the use of recurrence-quantification analysis (RQA) as opposed to cross-recurrence quantification analysis (CRQA) for the study of coordination dynamics in conversation. RQA is a more basic version of CRQA. It compares the same time series against itself in order to track cases of self-recurrence (points in time in which the same behaviour re-occurs within the same time series). For dialogues, this means creating a unique time series collapsing the speakers' speech streams into a unique conversational stream that, while respecting speaking turns, does not discriminate between speakers. The argument goes that RQA, unlike CRQA, would be able to capture not just matching or converging behaviour (i.e., interlocutors doing the same thing) but also complementary behaviour (i.e., interlocutors doing different things but in order to achieve a common underlying goal, such as two speakers asking and answering

\textbf{Figure 19} Example of windowed CRQA applied to a dialogue (whose silent turns were dummy coded). Window size = 20 and sampling interval = 5. Only the first three and the last two windows are displayed in the figure.
each other’s questions). This appears to be an intriguing and potentially promising approach, although confirmatory experimental or simulation studies still have to be conducted in this regard. Nonetheless, if applied to the investigation of word-level lexical convergence, RQA “dangerously” confounds self-repetition and interpersonal-repetition (i.e., convergence) patterns (see Figure 20 for an example). For this reason, we did not pursue this approach in our study. Note that we are not claiming that self-repetition is of no interest to the study of conversation. According to the interactive alignment model (Pickering & Garrod, 2004), for instance, the co-representation of linguistic information between the comprehension and the production systems, together with priming mechanisms, should lead to both between- and within-speaker convergence. The researcher must though be able to distinguish between interpersonal convergence and self-priming in their measurements.

Figure 20 Recurrence Plot (RP) of the conversation in Dialogue 1A. The conversation was coded as a unique conversational stream (not representing who is saying what). Circled recurrence dots represent cases of speakers 'converging' with themselves. Recurrence was limited to only one half of the RP as the two triangular areas capture exactly the same information. 
\%RR = 2.95, \%DET = 58.82, maxL = 5, L = 3.33, ENTR = 1.10, rENTR = 1
4.3 APPLICATIONS OF CRQA TO THE STUDY OF INTERPERSONAL VERBAL INTERACTION

CRQA has recently gained ground as a suitable statistical tool to quantify interpersonal coupling and linguistic convergence dynamics between interacting individuals (see Marwan, 2008). In this section, we review the main studies that have use CRQA for the study of conversation, and outline both the potential for, and the obstacles to, its application to linguistic data.

4.3.1 CRQA and conversation: the variety of non-linguistic coordination

At first, several studies used CRQA to explore how talking and interacting with another person could lead to forms of unintentional (i.e., undeliberate) behavioural coordination, suggesting that interactive language use may work as a medium for motor and attentional coupling (at least in cooperative contexts, see Paxton & Dale, 2013). This evidence shows the flexibility of CRQA, which has been successfully applied to a variety of time-series data: heart rate (Fusaroli, Bjørndahl, Roepstorff, & Tylén, 2015), skin conductance (Mønster, Håkonsson, Eskildsen, & Wallot, 2016), eye-movements (Richardson, Dale, & Kirkham, 2007; Richardson et al., 2009), postural sway (Shockley, Santana, & Fowler, 2003; Shockley, Baker, Richardson, & Fowler, 2007), and facial muscles (Louwerse, Dale, Bard, & Jeuniaux, 2012). As important as these precursor studies are, however, they measured the synchronization between continuous (in variability and time) behaviours; that is, time series that more closely resemble the temporal and coupling processes for which CRQA was originally developed but that differ substantially from the discrete (in variability and time) linguistic data that we are concerned with in this study.

4.3.2 Linguistic coordination: Lab-constrained and natural conversations

More relevant to the current study is the use of categorical CRQA for the quantification of linguistic coordination in experimentally controlled (e.g., Fusaroli & Tyléen, 2016) and natural conversations (e.g., Angus, Smith, & Wiles, 2012a). Research so far has focused on convergence at syntactic (repetitions of word-class n-grams; Dale & Spivey, 2006), morphemic (repetitions of three-character strings; Orsucci et al., 2006), lexical (repetitions of lexical bigrams; Fernández & Grimm, 2014) and prosodic (fundamental frequency; Fusaroli
& Tyléen, 2016) levels. Some studies went beyond form-based similarity and tweaked the tool to capture the degree with which, across utterances, interlocutors engaged in the same conversational topic (Angus, Smith, & Wiles, 2012b). Studies have mostly rely on the CRQA measure of %RR, which has been estimated and compared between adult-adult and child-adult dialogues (e.g., Fernández & Grimm, 2014), between informal and clinical conversations (Orsucci et al., 2013), and in longitudinal corpora (Cox & van Dijk, 2013). However, most of this evidence comes from exploratory or case-based studies (which prevented statistical comparison, e.g., Franco Orsucci et al., 2006), or from re-analysis of conversations produced in previous experiments (e.g., Fusaroli & Tylén, 2016; Gorman, Cooke, Amazeen, & Fouse, 2012; Mitkidis, McGraw, Roepstorff, & Wallot, 2015).

Dale and Spivey (2006) were the first to apply categorical CRQA to linguistic coordination (as shared production of syntactic bi-grams) in child-adult natural conversations. The study is important in that it shows how CRPs and CRQA can be employed to quantify different types of linguistic convergence while being sensitive enough to capture variabilities due to individual differences (e.g., speech proficiency) and context (see also Warlaumont et al., 2010 for similar operationalisations). The crude measure of cross-recurrence for the whole CRP (“global” %RR) was taken as a marker of the overall level of convergence between a child's and adult's syntactic production. The authors then estimated %RR only for a smaller band around the line of synchronization in the CRP. By construction (see section 4.2), this captures interlocutors' behaviour produced around the same time in the conversation and can be interpreted as short-term convergence. Lastly, to look at leading-following dynamics in the production of syntactic structures, %RR for the upper and lower triangular areas in the CRP was estimated separately (e.g., upper area = child is repeating/following the adult). Compared to chance, children showed higher syntactic coordination with their adult interlocutor both overall and between close conversational turns. Leader-follower dynamics were found to depend on the developmental stage of the child (with verbally older children converging less and leading more).

Importantly, all these studies show that CRQA can measure coupling dynamics not only across time series of continuous data that unfold over time simultaneously (e.g., postural sway) but also to multinomial time series that unfold in alternation (e.g., speech production
in conversation). Furthermore, CRQA can quantify not just the convergence/entrainment of behaviours (i.e., people producing the same behaviour with more or less tight temporal synchrony) but also their complementary coordination (i.e., people producing different behaviours to achieve a joint goal). One linguistic example is the investigation of turn-taking dynamics. Turn-taking behaviour has been operationalised in various ways: as nominal sequences of 1s (speaker A talking in a given time segment) and 0s (speaker A silent) (e.g., Warlaumont et al., 2010), as sequences of codes representing which interlocutor was talking (Gorman, et al., 2012), or as utterance lengths (0s for one-word utterances, 1s for two/three-word, and 2s for four-or-more-word utterances; Cox & van Dijk, 2013). Higher %RR or %DET have been interpreted as smoother exchanges (e.g., Reuzel et al., 2013) or dynamical rigidity (e.g., Gorman et al., 2012).

4.3.3 Statistical testing

Most CRQA studies have measured the coupling dynamics of interactions by comparing CRQA measures of real interactions against chance. What was this “chance” and why is it important?

Some studies have shuffled the order of the data within each time series (e.g., Louwerse et al., 2012). This removes the temporal dependencies in the data but maintains the distributional properties of the original series. It thus tests the hypothesis that the observed recurrence patterns in the conversation are due to the temporal (i.e., sequential) structure and not accidental distribution of the data (e.g., words; Gorman, Amazeen, & Cooke, 2010; Theiler, Eubank, Longtin, Galdrikian, & Doyne Farmer, 1992). Applied to lexical time series, individual speakers' word production would be turned into random strings of words. Given turn-taking, words should be shuffled across turns as within-turn randomization would not remove the temporal dependencies across speakers' utterances. A preferred alternative would be to shuffle one of the interlocutors' turn order (e.g., Healey, et al., 2014). This would dismantle temporal dependencies while maintaining distributional properties. However, the approach does not test directly the hypothesis that convergence is due to the coupling dynamics between the interlocutors, as these are removed together with the temporal dimension. That is, especially in strictly-structured task in which participants follow the same routine across dyads, a significance difference between the empirical and the shuffled
time series would not be able to clarify whether the found “above-chance” convergence is due to the temporal unfolding of the task or, rather, to interpersonal adaptation (as both were removed in the shuffled data).

Other studies have constructed pseudo-interactions obtained by shuffling pair arrangements (i.e., speaker 1 from dyad A is paired up with speaker 2 from dyad B; e.g., Dale & Spivey, 2006; Shockley, Baker, Richardson, & Fowler, 2007). Pseudo-interactions have the advantage of maintaining the temporal unfolding of the original interaction (and so of the experimental task) while disrupting both the distributional and the actual interpersonal coupling dynamics. Importantly, however, when the conversation is more strictly-structured and task-oriented (e.g., conversations of dyads assigned to the same experimental list or topic), this chance baseline is likely to preserve part of that between-subject linguistic convergence which is due to linguistic convention or task-demand, so distributional properties should be roughly maintained. As a result, significant differences with the original time-series are likely to be due to interpersonal coordination. However, its application is problematic when the time series analysed involves turn-taking, as different dialogues have different numbers of turns and different alternation patterns (Fusaroli, et al., 2014). A way out is to pair conversations by number of turns and disregard “unmatched” turns. An alternative approach would be to replace one speaker's production with utterances chosen at random points from random other dialogues (Howes et al., 2010). However, this baseline disrupts both temporal and interpersonal-adaptation dynamics and it is not clear what it would test for.

By dividing the empirical recurrence measures by the recurrence measures obtained for the “chance interactions”, some studies have obtained 'normalized' recurrence measures capturing above-chance coordination which has then been used for analysis (e.g., Warlaumont et al., 2010). Others have used as baseline the level of cross-recurrence of simultaneously-yet-individually-performing individuals (e.g., Ramenzoni, Riley, Shockley, & Baker, 2012). Fusaroli et al., (2015) combined the two approaches and had groups of participants performing a task together (joint) and individually but side-by-side (solo). Interpersonal behavioural synchrony was computed for each possible pairs within each group, and compared between the solo and the joint conditions, as well as between each possible
real pair and their pseudo-pairs. The clever set up was able to assess whether synchrony was due to simple co-presence (real vs. pseudo-pairs), task-demand (solo pairs vs. joint pairs, as only participants in the solo-task performed the same actions; participants in joint tasks likely performed complementary actions), or behavioural synchrony (real joint pairs vs. everything else).

As far as we know, few studies have so far compared the CRQA measures of competing theory-driven experimental conditions. When this has been done, most were re-analyses of data from previous experiments (e.g., Fusaroli & Tylén, 2016; Gorman et al., 2012; yet see Fusaroli et al., 2015), with the risk that effects could be confounded with other experimental manipulations (e.g., Mønster, et al., 2016).

4.3.4 Choice of CRQA measures, effects sizes and cross-interaction variability

Most of the findings reviewed above are based on the CRQA measure of “crude” cross-recurrence, with high $\%RR$ interpreted as evidence of coordination or alignment between interacting individuals. However, $\%RR$ has its limitation as a marker of entrainment at a linguistic level (see section 4.2). In general, CRP reflects rather sparse matrices when applied to linguistic data, given the many different things speakers can talk about, and the many different ways they can talk about them, and the time dynamics of turn-taking. For instance, $\%RR$ was considerably lower than 1% in Dale and Spivey (2006). When applied to conversations, other measures like $\%DET$ (as a marker of prolonged convergence) and $L$ or $maxL$ (as markers of over-time stable entrainment) are more meaningful and should be preferred (e.g., Fusaroli et al., 2014).

In most studies, effect sizes were not reported. Furthermore, the authors seldom discussed the statistical significance of the CRQA measures that were not reported (see Fusaroli et al., 2014 for recommendations in this regard). At the same time, several studies have emphasised the high variability in cross-recurrence across dyads and interactions. For instance, in Reuzel et al. (2013), staff-clients' coordination varied across interactions in the range of 17% - 70% for eye-gaze, and 42% - 70% for turn-taking (see also Cox & van Dijk, 2013; Shockley, et al., Folkwer, 2007 for comparable results).
In this chapter, we took a closer look at how categorical CRQA can be applied to nominal, noisy and stochastic dialogue data, and can capture convergence dynamics between interlocutors. We showed that CRQA provides a method to capture the overall/global degree of adaptation between two interlocutors' language use, and investigate leader-follower dynamics. We outlined the importance for modelling turn-taking of dummy-coding silence at the risk of mis-representing the actual convergence dynamics and inflating the cross-recurrence measures. When applied at word level, diagonal-based CRQA measures should be preferred over the crude cross-recurrence rate as they are more likely to track continued non-accidental convergence. Windowed CRQA allows quantifying shorter-term as well as the time-dependent patterns in speakers' convergence. Its implementation and results, however, depend on the non-straightforward choice of window size and sampling interval. In section 5.2.7.1, we will propose an iterative selection process based on maximising the recurrence-to-noise ratio. A question remains open. In the introduction to this thesis we distinguished between lexical repetitions between interlocutors that are driven by topicality, task demand and linguistic conventions, and those that are the result of communicative interpersonal convergence (i.e., driven by the other person's linguistic production). By itself, CRQA is not able to distinguish between these different cases. We suggested that, for more task-dependent and strictly-structured dialogues (e.g., lab-based ones) pseudo-interactions, in which a speaker's turns are interleaved with the turns of the speaker from a different conversation, could provide a suitable "chance" baseline to single out interpersonal coordination dynamics.
Chapter 5

SOCIAL IMPRESSION AND LEXICAL CONVERGENCE

5.1 THEORETICAL BACKGROUND

As we saw in Chapter 2, the language of people engaged in conversation with each other presents degrees of linguistic similarities that result from the mutual influence of each other's linguistic production. For instance, we are more likely to tell our friend that we are sorry she feels so unhappy if she sighed about how unhappy (rather than sad) she was. These effects have been explained in many different ways (e.g., Bock & Griffin, 2000; Brennan & Clark, 1996; Giles & Powesland, 1975; Fusaroli, et al., 2013; Pickering & Garrod, 2004; see section 2.1); in this chapter, we focus on the possible influence of social factors.

There is increasing interest in the effect of social factors on linguistic convergence (e.g., Wheatherholz et al., 2014). In this study we are concerned with what in Chapter 2 we defined as contextual social factors: characteristics of the situation in which the conversation unfolds that relate to how we perceive and manage interpersonal relationships; for instance, the interactional goal of the conversation and speakers’ perceptions about each other. Evidence from lexical convergence (Branigan et al., 2011; Genesee et al., 1995) has already shown that speakers' impressions about their interlocutor can impact the extent to which they converge on their lexical choices. However, most of these findings focused on beliefs about the other as a speaker (e.g., competent vs. non-competent; Branigan et al., 2011) and that can be explained in terms of audience-design (i.e., what the other is more likely to understand), and we are not addressing them here. Instead, we are interested in those beliefs that are based on the other's attitude and that reflect what we feel or think of the other as a “type of person”.

An increasing number of studies have started to look at the effect of social perceptions on linguistic convergence. They found that when speakers are positively impressed by another speaker, they are more likely to converge on this speaker’s verbal and linguistic style,
assumingly to express or even reinforce rapport (cf. section 2.1.2). For instance, the language of people who like each other shows a higher degree of similarities in terms of functional-word use (Gonzales et al., 2010), syntactic structures and positive/negative valence of their words (Moscoso del Prado Martin & Du Bois, 2015), and phonetic features (Pardo et al., 2012). However, as we discussed in Chapter 2, the correlational nature of most of these studies cannot clarify the directionality of the effect: did positive affect lead people to linguistically converge more, for instance to show affiliation, or did people who happened to linguistically converge more (for any other reasons, such as, perhaps, automatic interactive priming mechanisms; cf. Pickering & Garrod, 2004) develop a stronger relationship with each other? Recent studies have started testing these separate hypotheses. By experimentally manipulating the extent to which speakers may develop positive affect for their interlocutor, these studies then measured whether the developed interpersonal bond predicted speakers’ degree of linguistic convergence. For instance, Balcetis and Dale (2005) had participants do a picture-description task with a partner (in reality, a confederate) who, at the start of the session, came across as nice and well-disposed towards the participant or as arrogant and with a low consideration of them. Speakers converged syntactically more with the likeable partner than with the disagreeable one, thus confirming the link between positive affect and a high degree of linguistic convergence. However, in a second experiment, when the likeable versus arrogant attitudes were directed towards an inefficient experimenter, and thus could be interpreted as markers of annoyance (vs. understanding) towards the situation and not as gratuitous meanness, speakers revealed a reverse pattern and converged more with the annoyed than the sympathetic partner. These results are very interesting; they show that the path from speakers’ impressions about the other to linguistic convergence is not univocal (i.e., high degree of linguistic convergence being connected to likeable behaviour or other positive characteristics of the speaker). Indeed, linguistic convergence can be used as a compensatory strategy that helps speakers (re-)establish smooth interaction or deal with “difficult” interlocutors (see also Lev-Ari, 2015). Similarly, (Louwerse et al., 2012) found increased behavioural synchronization between participants involved in an interactive route-giving task, when the task and the communication got more challenging. These results extend independent evidence that linguistic convergence may be used by speakers as a means to meet social goals or manage social distance (e.g., Coyle & Kaschak, 2012). For instance,
linguistic convergence was used by member of an online community when providing support and showing empathy (Wang, et al., 2015), and speakers tended to diverge from other speakers whom they found socially/culturally antagonistic to mark disassociation (Bourhis & Giles, 1977).

Relevant to our study, the link between social perception and linguistic convergence seems to be mediated by the speaker's own attitude and dispositions, such that the characteristics of the interlocutor are not categorized passively but are “interpreted” in light of how we perceive ourselves and the situation. In other words, the effect of the other's attitude and behaviour on linguistic convergence may be modulated by how the speaker relates herself towards the displayed attitude. For instance, participants' re-use of the heard syntactic constructions in a non-interactive priming study was affected by how close they could relate to the ideo-political views of the person who produced the prime (Weatherholz et al., 2014). Abrego-Collier and colleagues (2011) found that participants converged more with the phonetic characteristic of a narrator toward whom they developed a positive disposition; a factor that reflected the combination of the narrator's intrinsic characteristics (e.g., sexual orientation), the content of the narrative (e.g., positive vs. negative) and the speaker's own attitude. Similarly, the likelihood with which participants in an intelligence test re-used the syntactic structures heard in the recorded voice-message of a previous participant depended on how smart the other was perceived to be compared to how smart the participants felt (Lev-Ari, 2015).

In sum, there seem to be two trends in the published literature. First, social perception, that is the perceived attitude and characteristics of our interlocutors, seems to affect our degree of linguistic convergence conditionally on how they make us feel about ourselves and relate to the other. Second, the link between others' attitudes and linguistic convergence is more complex than convergence being always associated with perceiving (or creating) positive affect and divergence with perceiving (or creating) disaffiliation. In contrast, linguistic convergence may be used as a compensatory tool to “win over” a difficult interlocutor.

Nonetheless, the reviewed studies on social perception and degree of syntactic convergence have primed speakers' social perception of the other in a passive manner: using a pre-registered voice message and tasks for which the perceived attitude and characteristics of the
other did not actually affect the speaker. Furthermore, linguistic convergence was not measured between the participant and the other person in interaction but in subsequent, unrelated and non-interactive tasks. While the absence of interaction (and expectation thereof) allowed the authors to investigate the social modulation of linguistic convergence beyond the communicative requirement of an interaction, it limits theoretical conclusions as to whether social perception may have made certain linguistic characteristics (e.g., one syntactic alternative) more salient and so more likely to be re-used in a subsequent separate task. Thus, while these findings can inform the extent to which social factors can automatically affect information processing (e.g., Weatherholz et al., 2014) and change one's linguistic representations in memory (e.g., Lev-Ari, 2015), they provide no evidence for communicative adaptation in terms of coordination with the speaker. Finally, most studies focused on syntactic or phonetic convergence which we saw present characteristics that do not make it directly comparable to lexical convergence (e.g., syntactic choices and phonetic characteristics are less related to meaning; cf. section 2.6.2).

Our study seeks to explore further the effect of social perception on linguistic convergence. Specifically, we aim to test the social accounts’ claim that linguistic convergence (in this case, lexical convergence) is linked to how (positively vs. negatively) speakers perceive the other and to the way speakers manage social dynamics. Importantly, we are not committing to the claim that these links, if found, are conscious and deliberate (see section 2.2).

We manipulated the social perception of the other by having participants engage in a decision-making game ostensibly together (in reality with a pre-set computer program). Importantly, in this kind of game the behaviour of the players affects each other's performance in terms of gains in the game. Hence, the attitude of the partner has a direct and tangible consequence on each participant. More specifically, we used the Public Goods Game (PGG), which has been extensively used to measure trust and cooperation (e.g., Mitkidis et al., 2015; Rand, Greene, & Nowak, 2012).

In the PGG, participants decide how much to contribute to a common investment which is then proportionally increased and divided equally between them. The game is set up so that the pair's collective interests are maximized when both participants contribute maximally. At an individual level, an individual’s interests are best pursued by contributing much less than
the partner. A safe strategy to avoid any losses is not to contribute at all, that is to defect (see Method section for more details). The game has, thus, a clear element of cooperative (and trusting) versus individualistic (and mistrusting) behaviour (Andreoni & Andreoni, 1995; Bos, Olson, Gergle, Olson, & Wright, 2002). In particular, investing collectivistly requires one to trust their partner will also invest collectivistly, and thus to act on the positive expectations that they have of the other (e.g., Riegelsberger, Sasse, & McCarthy, 2003; Scissors et al., 2008). In contrast, participants who defect show a lack of willingness to trust their partner and instead act on safe individualistic grounds.

In our study, participants either played against a (computer-simulated) “other” who showed trusting and collectivist or mis-trusting and individualistic behaviour. They then performed an independent discussion task, this time with the real other participant, where we measured their degree of lexical convergence. This set up allowed us to estimate whether participants’ impressions about the other, built through the recent game experience, affected their tendency to converge on each other’s lexical choices.

We hypothesised that, if social perception affects lexical convergence such that the perceived positive attitude and the extent to which we are able to relate to the other leads to more convergence, then we should find that participants playing against a trusting and collectivist “partner” should converge more with them in the subsequent task than participants who have played against the mistrusting and individualistic “partner”. This pattern could arise either as a result of the increased rapport developed with the collectivist co-player or increased social distance towards the individualistic co-player. We call this the social distance hypothesis. In its support, a previous study that looked at linguistic mimicry between players in an interactive social dilemma game found a positive correlation between the level of trust between pair members and measures of linguistic convergence (Scissors, et al., 2008). Also, the level of nonverbal mimicry by one negotiator has been found to increase the level of trust felt by the other, in a way that facilitated negotiation outcomes (Maddux, Mullen, & Galinsky, 2008).

However, if speakers mostly use linguistic convergence as a social repair tool when dealing with challenging interacting partners, as other studies have suggested, then we should find the opposite pattern of results: participants playing against the mistrusting individualistic
“partner” should show a higher propensity to converge than participants who have played against the trusting and trustworthy “partner”. We called this the linguistic convergence as compensatory strategy hypothesis.

A second theoretical issue we are concerned with is the interactive goal of the conversation. Previous research has found that the degree to which interlocutors linguistically converge with each other depended very much on the extent to which their conversation was directed to a common goal. For instance, interlocutors converged substantially more if working toward achieving a specific goal together (e.g., instructing each other to draw routes on a map) than when they were conversing freely (Reitter et al., 2006, 2014). Other studies have found no above-chance degree of linguistic convergence in open-ended spontaneous conversations (Healey, et al., 2014). In this study, we decided to measure lexical convergence in two separate-yet-related subtasks that differed in interactional goal. In the first subtask, participants discussed the usefulness for survival of a set of items. Even if the conversation had a purpose (i.e., a precise topic of discussion), it did not involve any joint decision-making process and participants had no specific goal to achieve. In the second subtask (unknown to participants while carrying out the first part), participants had to agree on which five items to carry away with them. Thus, the conversation had a clear joint decision-making element and common goal. Importantly, the required level of referential communication was the same across subtasks (i.e., both subtasks required participants to talk about the same fifteen items and linked them to survival). If the interactional nature of the conversation affects interpersonal linguistic convergence, as previous findings suggest, then we should observe greater linguistic convergence in the second goal-oriented subtask. We do not have any predictions regarding a possible interaction between social perception and nature of the task.

We measured lexical convergence at two levels:

- Word forms: All linguistic items were included in this measure in the form they were typed in by participants. The measure also includes those non-content expressions (e.g., function words) whose repetition across speakers has been found to correlate with social outcomes (e.g., Ireland et al., 2011; Richardson et al., 2014).
• Stems: This was a refined form of lexical convergence: focused on content words and independent of syntactic choices. Excluding function and no-content words, it includes more task-specific language, which has been shown to be more predictive of interaction outcomes and success (e.g., Fusaroli et al., 2012).

Finally, we also measured and controlled for those individual-difference traits that have been suggested to relate to a person's propensity to linguistically align (i.e., perceived interpersonal similarity, preference for compromise, and dominance during conflict; Weatherholtz et al., 2014).

5.2 EXPERIMENT 4

5.2.1 Methods

5.2.1.1 Participants

64 native speakers of British English from the University of Edinburgh's student community participated in the experiment (42 female, 22 male; mean age = 21.6 years, SD = 2.7 years) in return for £6.20. The participants were paired into 32 same-gender dyads; members of each dyad did not know each other beforehand.

20 native speakers of British English (12 females, 8 males; mean age = 21.2 years, SD = 4.8 years), paired into 10 same-gender dyads of unacquainted members participated in the pilot. They were paid £3.20.

5.2.1.2 Design

Dyads were randomly assigned to the collectivist or the individualistic Partner condition.

Figure 21 provides an overview of the three subsequent stages in the Experiment. Note that participants in the Pilot only took part in the Public Goods Game and then filled out a manipulation-check questionnaire that was not administered to the participants in the Experiment. This was because the manipulation-check could cue Experiment participants to the real purpose of the study, biasing the study's results.

28 The experimental design was co-ideated with Jaroslaw Lelonkiewicz.
5.2.1.3 The Public Good Game (PGG)

The PGG was our manipulation of social perception of the partner. Participants were told that, for this task, they would be paid according to how much they gained through their actions in the game (£0.01 for every second credit). Crucially, the participant's payoff depended on how much both she and her partner behaved in the game (see below for more details). The maximum possible payoff was £3.20. Dyad members took part in the PGG thinking that they were playing with each other; in reality, they individually played with a pre-set computer programme. The game consisted of 10 single-shot trials. In each trial, each participant was given an endowment of 40 credits and decided, unbeknownst to the other player (in reality, the computer), how many of these credits to keep vs. contribute to a public pot (the partner-computer was programmed with a pre-set range of values to contribute; see below). The credits in the public pot were then multiplied by a factor of 1.2 and divided equally between the two. In each turn, the participant’s payoff was made of half the credits that were invested in the pot (after multiplication) summed to the credits that she had not invested (e.g., in the example in Figure 22, the participant invests 20 of her 40 credits for that turn. Her payoff is 53: 20, i.e., the credits she did not invest, plus 33, i.e., half of the credits...
in the pot). Hence crucial to our manipulation, the participant's payoff depended on how much both she and her “partner” had invested in the public pot. Specifically, participants would obtain higher gains if they invested little and their partner invested more; their highest gains could be obtained if they contributed zero and their partner contributed all their endowment. Acting collaboratively could also lead to gains but only if the other person acted collaboratively too. That is, acting collaboratively was a more risky behaviour that could lead to big losses and thus required a higher level of trust in the other. A safe behaviour would be to contribute none or little to preserve at least one's original credits.

After each turn, the participant saw a summary of both “participants’” contributions to the pot, their payoffs for that turn and how much they had gained (or lost) with respect to their initial 40 credits (see Figure 22 for an example trial). In this way, participants received indirect feedback on each turn about their partner’s individualistic or collectivist behaviour. At the end of the 10 turns, participants were informed how many credits they had collected in total and the corresponding payment amount in British pounds.

*Figure 22 Example trial (collectivist Partner condition).*
Partner (computer)'s behaviour in the two experimental conditions

The algorithm for the partner-computer’s behaviour in the two experimental conditions is reported in Appendix I and J. With the exception of turn 1, the partner-computer’s contribution at turn \( n \) was based on the participant’s contribution at turn \( n-1 \). The partner-computer’s behaviour was coded in E-Basic. The experiment was run in E-Prime.

Collectivist

The collectivist partner-computer started the game as a trusting contributor, investing 35 credits into the public pot (turn 1). In all the subsequent trials, the computer showed what could be defined as conditional cooperation with a touch of warm-glow (Fischbacher & Gächter, 2010; Gächter, 2007; Szolnoki & Perc, 2012). The partner-computer always contributed slightly more than the amount the participant contributed in the preceding turn (with the exception of when the participant had contributed above 30, in which case the partner-computer could contribute a higher or equal amount). Importantly, if the participant was not giving as much as the computer did, the computer tended to reduce the amount it shared in the next turn while maintaining a cooperation-seeking or warm-glow attitude by still contributing slightly more than the participant had done (see Table 8, for an example).

### Table 8

Example game (Collectivist Partner). Payoff indicates the total number of credits collected in a given trial (number of credits gained or lost with respect to the original 40).

<table>
<thead>
<tr>
<th>Trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer:</td>
<td>35</td>
<td>26</td>
<td>35</td>
<td>34</td>
<td>38</td>
<td>35</td>
<td>39</td>
<td>38</td>
<td>38</td>
<td>40</td>
</tr>
<tr>
<td>Subject:</td>
<td>20</td>
<td>30</td>
<td>30</td>
<td>35</td>
<td>35</td>
<td>37</td>
<td>36</td>
<td>38</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td><strong>Payoff</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer:</td>
<td>38</td>
<td>47.6</td>
<td>44.0</td>
<td>47.5</td>
<td>45.3</td>
<td>48.2</td>
<td>46</td>
<td>47.6</td>
<td>48.2</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>(-2.0)</td>
<td>(+7.6)</td>
<td>(+4.0)</td>
<td>(+7.4)</td>
<td>(+5.8)</td>
<td>(+8.2)</td>
<td>(+6.0)</td>
<td>(+7.6)</td>
<td>(+8.2)</td>
<td>(+8.0)</td>
</tr>
<tr>
<td>Subject:</td>
<td>53</td>
<td>43.6</td>
<td>49</td>
<td>46.3</td>
<td>48.8</td>
<td>46.3</td>
<td>49</td>
<td>47.6</td>
<td>47.2</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>(+13.0)</td>
<td>(+3.6)</td>
<td>(+9.0)</td>
<td>(+6.3)</td>
<td>(+8.8)</td>
<td>(+6.3)</td>
<td>(+9.0)</td>
<td>(+7.6)</td>
<td>(+7.2)</td>
<td>(+8.0)</td>
</tr>
</tbody>
</table>
The individualistic partner-computer started the game as a mistrusting contributor, investing 15 credits into the public pot (turn 1). In all the subsequent trials, the partner-computer behaved like a defector, that is, someone who believes that the other will not be contributing cooperatively and actually contributes nothing or very little himself. Thus, the individualistic partner-computer never looked for positive reciprocal adjustment; it always contributed less than the amount that the participant had contributed in the previous trial (unless this was two zeros in a row, in which case the partner-computer invested zero), and never more than 15 credits (see Table 9, for an example).

### Table 9

<table>
<thead>
<tr>
<th>Trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer:</td>
<td>15</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Subject:</td>
<td>20</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Payoff</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer:</td>
<td>46</td>
<td>44.8</td>
<td>39.8</td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>(6.0)</td>
<td>(+4.8)</td>
<td>(-0.2)</td>
<td>(+3.0)</td>
<td>(+3.0)</td>
<td>(+3.0)</td>
<td>(=)</td>
<td>(=)</td>
<td>(=)</td>
<td>(=)</td>
<td>(=)</td>
</tr>
<tr>
<td>Subject:</td>
<td>41</td>
<td>37.8</td>
<td>40.8</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>(+1.0)</td>
<td>(-2.2)</td>
<td>(+0.8)</td>
<td>(-2.0)</td>
<td>(-2.0)</td>
<td>(-2.0)</td>
<td>(=)</td>
<td>(=)</td>
<td>(=)</td>
<td>(=)</td>
<td>(=)</td>
</tr>
</tbody>
</table>

### 5.2.1.4 Dialogue (plane-crash survival) task

Dyad members engaged in a timed conversation, this time with the real other. The conversation task was used to measure lexical convergence. During the task, participants were seated in separate lab booths and communicated with each other using the DiET chat-
tool (Mills & Healey, submitted), a text-based chat software that automatically transcribes all key presses. At the beginning, participants were given 5 minutes to individually read the instructions and familiarise themselves with the following scenario (see Appendix K for more details). A pdf with photos of the 15 items was simultaneously displayed on their screens (see Appendix L).

Plane-crash scenario

The participants were told that they were two work colleagues flying together to Beijing. They had just crash-landed in the woods of the Ural Mountains (Russia). It was 11:32 AM in mid-January. They were the only survivors. Neither of them was seriously injured. A last communication from the captain had informed them that the plane was trying to avoid a storm and it was thus considerably off course: roughly 120 miles southwest of a small town that was the nearest known habitation. They were in a wilderness area made up of thick woods broken by many lakes and rivers. The last weather report indicated that the temperature would reach minus fifteen degrees centigrade in the daytime and minus twenty-five at night. They were dressed in winter clothing appropriate for city wear, and there was no phone signal or internet coverage in the area. Luckily, they had found two portable two-way radio transceivers in the jackets of two Russian soldiers who had died on impact. The transceivers could only be used to send typed messages to each other, so they took one receiver each and went to explore the crashed aircraft.

The participants’ task was to inspect the aircraft with the aim to gain an understanding of the situation with their partner. They had 15 minutes to discuss the usefulness for survival of each of the items remaining intact on the aircraft. They were informed 10 minutes, 5 minutes and then again 2 minutes before the time ran out. At the end of the 15 minutes, participants were told by the experimenter about the extra surprise task (see Appendix M). The plane wreckage was about to blow up and they had 5 minutes to decide which 5 items to take away with them. They were also informed that there was a correct solution to this task, prepared by U.S. Army Instructors on Wilderness Survival. As in typical online chat programs, participants could type at the same time but could not access what the other was typing until the message was sent out (they could see “[name] is typing” while the other was typing). Unlike typical online chat programs, however, the history of the conversation was not
available to participants and only the last message sent remained visible on screen. As we were interested in lexical convergence (i.e., the extent to which participants’ lexical choices mimic each other over time), this was done to avoid the effect being confounded with participants simply reading off the words from their screens.

5.2.1.5 Individual Differences

We measured participants’ Perceived Interpersonal Similarity (PIS), Preference for Compromise (PfC), and Dominance during Conflict (DdC). We chose these three measures as they were found to influence the degree of syntactic convergence in a previous study (Weatherholtz, et al., 2014). The measures consisted of 10 Likert-scale survey items, phrased as statements; participants indicated the extent to which they agreed with each statement from 1 (strong disagreement) to 7 (strong agreement). The full set of Likert-scale items is available in Appendix N. The questionnaire also collected demographic information (i.e., gender, age, spoken British English dialect, degree, and the number of psychology experiments the participants had previously taken part in). The individual difference items were mixed in with filler items randomly taken from an alcohol questionnaire and personality questionnaire (Leeds Student Medical Practice, 2009; Francis, Lewis, & Ziebertz, 2006). Participants were also asked what they thought the purpose of each experiment was and whether they had any comments on either.

5.2.1.6 Procedure (Experiment)

The PGG and the Dialogue survival task were advertised and presented as separate stand-alone experiments that were administered in one single session. This was done to minimize the risk of participants inferring that the PGG was used to affect their behaviour in the subsequent interaction.

Each participant dyad arrived to the lab. The experimenter welcomed the participants making sure to mention both their names, and presented herself. She then briefly explained the nature of the two experiments and highlighted that, for the PGG, participants would be paid according to how much they gained in the game itself. Participants were then randomly assigned to one of two sound-proof lab booths. Participants read the information and signed the consent form for the first experiment (the PGG) and were then invited to read the
instructions of the game from their computer screens. The instructions asked them to type in their name so that it could be displayed on their partner's computer during the game (in reality, the partner's name was pre-inserted by the experimenter). The rules of the game were explained and a concrete example was provided (see Appendix O). At the end of the instructions, participants were asked to call the experimenter. This was done to answer participants' questions but also to increase participants' belief that they would be playing the game with each other. The first participant to finish reading the instructions was in fact asked to wait for the other before pressing start. When both participants were ready, they were told to start. At the end of the PGG, participants were given a one minute break to prevent them from realising that they had finished the game at different times (if that happened). Participants then signed the consent form for the second and last experiment (the dialogue survival task). After the dialogue task, participants filled out the post-interaction survey. They were then debriefed and paid. They were paid the maximum possible payoff for the PGG and explained that they had actually played with a pre-set computer programme. The experimenter was blind to the Partner manipulation throughout the whole experimental session.

5.2.1.7 Procedure (Pilot)

The procedure was as in the Experiment with the following difference. At the end of the PGG, participants were asked to fill out a manipulation-check questionnaire. We asked participants to reflect about the interaction they had just had with the other person during the PGG and their partner's behaviour (in reality, the computer's). The measures consisted of nine 7-point Likert-scale items, phrased as statements, adapted from Byrne (1971), Fraley and Aron (2004), and Treger, Sprecher, and Erber (2013) (see Appendix P). Participants were also asked whether and how their partner's playing attitude influenced their own, and whether they had any comments on the experiment or the interaction. Finally, they were explicitly asked whether they thought they had played with a person or with a simulated partner, and, if so, the extent to which they thought the computer's behaviour was human-like. Participants were then debriefed and paid the game's maximum possible payoff.
5.2.2 Pilot (manipulation check)

5.2.2.1 Analysis and Results

Data from the two members of each dyad were analysed independently.

**Manipulation-check 1: Playing with the real or a simulated partner?**

Of the 20 participants, 19 (95%) affirmed that their contributing behaviour was influenced by the other person's behaviour (in reality, the computer).\(^29\)

No participant spontaneously mentioned any concerns about the nature of their playing partner in the open-comment session. However, when they were explicitly asked about it, 16 participants (80%) thought they interacted with a person and 4 with a simulated partner. Of the four participants who responded 'simulated partner' (2 female, 2 male), 3 were in the collectivist and 1 in the individualistic Partner condition. A Fisher's Exact test revealed no evidence that either Partner condition led to more 'simulated partner' responses (\(p = 0.582,\) odds ratio = 0.28, 95% CI [0.005; 4.35]). All four participants judged the simulated partner's behaviour to be either fairly human-like or human-like.

**Manipulation-check 2: Social perception of the other participant**

Responses to the five Likert-scale items were numerically coded with a value from 1 to 7. Table 10 reports median and interquartile range for the two Partner conditions.

---

\(^{29}\)The participant declaring their behaviour had been unaffected by the other person's behaviour was in the Individualistic Partner condition, and replied “person” to the question whether the other was a real person or a simulated partner.
Table 10

<table>
<thead>
<tr>
<th>The other person's behaviour impressed me as being (1 = extremely unfair) – (7 = extremely fair)</th>
<th>Collectivist</th>
<th>Individualistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel that I would probably (1 = dislike) – (7 = like) this person very much</td>
<td>6 (1.00)</td>
<td>3 (3.75)</td>
</tr>
<tr>
<td>The other person came across as being self-centred (1 = not at all) – (7 = very much so)</td>
<td>2 (0.75)</td>
<td>5.5 (2.75)</td>
</tr>
<tr>
<td>The other person came across as being trustworthy (1 = not at all) – (7 = very much so)</td>
<td>5 (1.5)</td>
<td>1.5 (2.5)</td>
</tr>
<tr>
<td>I feel that I would (1 = very much dislike) – (7 = very much enjoy) working with this person on a future task</td>
<td>5 (0.75)</td>
<td>3 (1.75)</td>
</tr>
</tbody>
</table>

We treated the data as ordinal and conducted separate permutation-based Mann-Whitney tests, one on each survey item, using the \textit{wilcox.test} function in the R \{coin\} package (Hothorn, Hornik, van de Wiel, Zeileis, 2008). We went for the permutation approach due to ties in the data. To control for multiple comparisons, p-values were adjusted following the Benjamini and Hochberg's False Discovery Rate procedure. All reported tests were one-tailed.

People in the Collectivist condition judged their partner's behaviour to be more fair ($Z = 3.20$, $p = .0026$, $r = .71$), and their partner to be more trustworthy ($Z = 2.72$, $p = 0.0027$, $r = .61$) and less self-centred ($Z = -2.38$, $p = 0.0082$, $r = -.53$) than people in the Individualistic condition. They also liked their partner more ($Z = 2.87$, $p = 0.0027$, $r = .64$) and were more willing to work with them in the future ($Z = 2.78$, $p = 0.0027$, $r = .62$).\footnote{We re-ran the analysis without the four participants who reported believing themselves to have played with a simulated partner. All significant results were confirmed.}

5.2.2.2 Discussion

The results of the pilot confirmed that our experimental manipulation worked as expected. First, the playing behaviour of the computer was believably human; the majority of our participants did indeed believe that they had played the game with the real other participant. The two Partner conditions were equally effective in this regard. Second, participants judged...
the playing partner differently in the two conditions. In line with our expectations, participants in the collectivist condition liked their partner more and judged them as more trustworthy, fairer and less self-centred. On the ground of these results, we took the PGG forward to the proper Experiment.

5.2.3 Lexical data: Coding

5.2.3.1 Word-form

This analysis was based on the transcripts at word level of all the chat conversations. We constructed a series of regular expressions to automatically parse the transcripts and prepare them for analysis. We transformed every word to lower case and removed all punctuation (so that “What”, “what?” and “what” counted as the same word) with the exception of the apostrophe. We kept in emoticons (e.g., “:-D”) and digits; all symbols that were used as abbreviation for a word (e.g., “=” for “equals”, “&” for “and”) were also kept in and left as symbols. We left misspelled words in their misspelled form. As in common chat-tools, participants had the chance to correct spelling mistakes before posting their message, and this is what they usually did. We assumed that correcting such spelling mistakes would have artificially inflated recurrence rate (i.e., lexical convergence).

The cleaned texts were then entered in an automated procedure that transformed each transcript into two discrete sequences of (numeric) codes, with one unique code for word form, corresponding to the sequences of words produced by the two interlocutors (see example below). This numeric sequences respected both the order with which words were produced and the temporal structure of turn taking (see section 4.3.3 for details). The choice of modelling silence was made, first, to accurately represent the temporal constraints of our chat design: turn-taking was locked and only one participant at a time could post their message. Second, failing to model silence would artificially inflate cross-recurrence and, hence, CRQA measures (see section 4.2.1).

The two procedures were programmed and run in R.

(Original transcript)

SpeakerA: I think the knife rather than the gun?
Speaker B: Yes me too. I think it would be more useful
Speaker A: I think the torch could be better though than either, we can collect berries and stuff rather than hunt meat?
Speaker B: the knife might be useful for other things though

(Cleaned text with dummy-coded silence)

Speaker A: i think the knife rather than the gun 999 999 999 999 999 999 999 999 i think the torch could be better though than either we can collect berries and stuff rather than hunt meat
Speaker B: 888 888 888 888 888 888 888 yes me too i think it would be more useful 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 the knife might be useful for other things though

(Final numeric sequences)

Speaker A: 12 25 23 14 20 22 23 10 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999 999
Speaker B: 888 888 888 888 888 888 32 15 27 12 25 13 31 2 18 29 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888

5.2.3.2 Stems

To extract stemmed words from the chat conversations, we applied the Porter stemming algorithm to the transcripts using the {SnowballC} package in R (version 0.5.1; Bouchet-Valat, 2014). We first corrected all spelling mistakes, and removed punctuation, digits, symbols and stop words (i.e., the most common content-poor words like articles, determiners, quantifiers, etc.) using the text mining R package {tm} (version 0.6-2; Feinerer & Hornik, 2015). We manually checked the stemmed transcripts and corrected eventual mistakes (e.g., we coded the different tenses of the same modal verb as the same verb instance, so that “could”, “cannot”, “can”, “couldn’t” were the same stem “can”).

The transcripts were then entered in an automated procedure that transformed each transcript into two discrete sequences of mutually exclusive (numeric) codes corresponding to the two interlocutors’ speech behaviours at stemmed-word level. The procedure was identical to the one reported in the previous section.
(Stem-based transcript)

SpeakerA: think knife rather gun
SpeakerB: think will use
SpeakerA: think torch can better though either can collect berri stuff rather hunt meat
SpeakerB: knife may use thing though

(Final numeric sequences)

SpeakerA: 14 8 11 6 999 999 14 16 2 15 5 3 4 1 12 11 7 10 999 999 999
SpeakerB: 888 888 888 888 14 18 17 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888 888

5.2.4 Individual Difference Measures (IDs) and Public Good Game

Participants’ responses were summed to produce their final perceived interpersonal similarity score (PIS; Cronbach's $\alpha = 0.83$), tendency to dominate during conflict score (DdC; Cronbach's $\alpha = 0.81$) and preference for compromise score (PfC; Cronbach's $\alpha = 0.85$), with higher scores indicating, respectively, higher levels of perceived interpersonal similarity, a stronger tendency to dominate during conflict and a stronger preference for compromise (summary statistics are reported in Table 11).

Table 11

<table>
<thead>
<tr>
<th></th>
<th>Collectivist (M)</th>
<th>Individualistic (M)</th>
<th>pooled SD</th>
<th>Cohen's d</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIS</td>
<td>19.97</td>
<td>18.66</td>
<td>4.48</td>
<td>0.29</td>
</tr>
<tr>
<td>PfC</td>
<td>12.97</td>
<td>14.81</td>
<td>3.85</td>
<td>-0.48</td>
</tr>
<tr>
<td>DdC</td>
<td>9.84</td>
<td>10.31</td>
<td>3.92</td>
<td>-0.12</td>
</tr>
</tbody>
</table>

The possible score range was 3-28 for PIS, and 3-21 for PfC and DdC.

As the final analysis was at the dyad level, we used the mean scores of PIS, DdC and PfC for each dyad to assess the overall level of that measure for the dyad (e.g., were the two
individuals *together* high or low in DdC?). We also calculated the absolute value of the difference in scores for each dyad to test for effects of *similarity* (or *non-similarity*) in PIS, DdC and PfC between dyad members (e.g., were the two individuals similarly high in DdC?). Table 12 reports standard deviation, range and correlations of the three IDs. Means and absolute differences were z-score transformed before being included in any analysis.

**Table 12**

Pearson correlations between mean and difference scores of the three IDs. Standard Deviation and [observed range] on diagonal.

<table>
<thead>
<tr>
<th></th>
<th>mean PIS</th>
<th>diff PIS</th>
<th>mean PfC</th>
<th>diff PfC</th>
<th>mean DdC</th>
<th>diff DdC</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean PIS</td>
<td>4.13 [6-24]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>diff PIS</td>
<td></td>
<td>0.39 [0-8]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean PfC</td>
<td>0.18 [7.5-19.5]</td>
<td>0.11 [0-11]</td>
<td>2.88 [6-15.5]</td>
<td>-0.14 [0-11]</td>
<td>0.48 [6-15.5]</td>
<td>0.25 [1-15]</td>
</tr>
<tr>
<td>diff PfC</td>
<td>-0.14 [0-11]</td>
<td>-0.27 [0-8]</td>
<td>-0.16 [0-11]</td>
<td>2.78 [0-11]</td>
<td>-0.06 [6-15]</td>
<td>-0.25 [1-15]</td>
</tr>
<tr>
<td>mean DdC</td>
<td>-0.12 [6-15.5]</td>
<td>-0.36 [0-11]</td>
<td>0.04 [6-15.5]</td>
<td>0.48 [6-15.5]</td>
<td>0.25 [1-15]</td>
<td>0.25 [1-15]</td>
</tr>
</tbody>
</table>

No correlation was significant (*p* values adjusted by Holm correction for multiple tests)

Finally, the amount contributed by participants in the first trial of the Public Goods Game, when they were unaware of their “partner’s” playing attitude, did not differ between conditions (Collectivist M = 22.34, SD = 13.09; Individualistic M = 21.16, SD = 9.15; *t*(55.5) = 0.42, *p* > .5, Cohen's *d* = .11, 95% CI [-.40; .62]). This confirms that participants had a similar initial approach to the game across conditions, and that any difference in their subsequent behaviour was due to our experimental manipulation.31

31 A 2x2 mixed ANOVA on the amount contributed by participants found indeed a significant Partner by Trial (1 vs. 10) interaction *F*(1,62) = 83.51, *p* < .0001. Post-hoc t-tests in each Partner condition found that participants in the Individualistic group contributed substantially less in trial 10 compared to trial 1 (M_trial1 = 21.16, M_trial10 = 2.97, *t*(45.4) = 10.07, *p* < .0001, Cohen's *d* = .74, 95% CI [.38; 1.11]). Participants in the Collectivist group contributed more in turn 10 vs. turn 1 but the difference did not reach significance (M_trial1 = 22.34, M_trial10 = 27.50, *t*(61.9) = -1.56, *p* = .125, Cohen's *d* = 1.94, 95% CI [1.51; 2.37]).
5.2.5 Recap of Predictions

Let us remind ourselves what the CRQA measures mean for linguistic data. \( DET \) represents the magnitude of lexical convergence. \( rENTR \) captures how predictable a conversation is in terms of the variability in the patterns of repeated words; the more predictable a conversation is, the more regular speakers’ patterns of convergence and so the lower the \( rENTR \). Both \( L \) and \( maxL \) track the stability of the conversation; in particular, conversations that contain lots of linguistic routines should show high \( L \). We will not consider \( maxL \), since its interpretation is unclear in the context of linguistic convergence. A dialogue could in fact display a very high \( maxL \), and so being misleadingly interpreted as highly stable, in the extreme case when interlocutors only repeated a very long sentence but no other word between each other in the whole conversation.

According to the social distance hypothesis of linguistic convergence, we should find a greater extent of (higher \( DET \)) and more stable (higher \( L \)) lexical convergence, and more predictable patterns of conversation (lower \( rENTR \)), in the Collectivist compared to the Individualistic group.

In contrast, according to the compensatory strategy hypothesis, we should find higher levels of \( DET \) and \( L \), and a lower level of \( rENTR \) in the Individualistic compared to the Collectivist group.

Finally, if the interactional goal of the task affects lexical convergence, we should find higher levels of \( DET \), and \( L \), and a lower level of \( rENTR \) when the task entails a precise joint decision-making element (5min task) compared to when it does not (15min task).

5.2.6 Analysis and Results

5.2.6.1 General note to the analysis

If the null-hypothesis significance testing results were “on the edge” (i.e., \( 0.02 \leq p\text{-value} < 0.15 \)), we decided to be conservative and verify the results by estimating the Bayes Factor (BF). BF calculates the empirical evidence in favour of the alternative hypothesis (e.g., the effect size for the difference in \( DET \) between Partner conditions is not 0) as opposed to the null hypothesis (i.e., the effect size is 0), or vice-versa. We estimated two types of BF. For
the regression coefficients in mixed-effects models, we used the BIC approximation of BF (Masson, 2011; Wagenmakers, 2007). For independent two-sampled permutation-test results, we used the testBF function in the {BayesFactor} package in R (Morey & Rouder, 2015; Rouder, Speckman, Sun, Morey, & Iverson, 2009); we conducted the BF with the default prior (Cauchy r = .71) and then performed a sensitivity analysis by varying r to represent wide (r = 1), ultra-wide (r = 1.41) and thin (r = .5) prior distributions of effect sizes (Morey, 2014, February 23; Schönbrodt, 2014, February 21). A BF higher than 3 constitutes convincing evidence in favour of the assessed hypothesis (Kass & Raftery, 1995).

5.2.6.2 Word-form lexical convergence

Models and analyses specification

To examine whether the extent, the predictability and stability of the word-form convergence between interlocutors differed in the two Partner conditions, and whether it was modulated by task demand, we conducted a series of mixed-effects regression analyses on the CRQA measures of DET, rENTR, and L. All analyses included Partner (Collectivist = .5, Individualistic = -.5), Task (15min = .5, 5min = -.5) and their interaction as fixed-effects predictors (unless otherwise specified), and the maximum random structure justified by the design (i.e., a by-dyad random intercept; Barr et al., 2013). Observed and model-predicted mean values are shown in Figure 23.

Both DET and rENTR were continuous proportion data which made beta (mixed-effects) regression (Ferrari & Cribari-Neto, 2004; Ospina & Ferrari, 2012) a suitable modelling choice given this bounded nature of the outcome variables. In the regression model for rENTR, Task was further centred to account for the imbalance in the dataset (15min = .31, 5min = -.69), due to the many conversations in the 5min task that showed no variation in diagonal length (i.e., interlocutors always and only converged on maximum two words in a row) and for which rENTR could not be calculated. Both models were implemented using the {gamlss} package in R (Rigby & Stasinopoulos, 2005; see also Stasinopoulos & Rigby, 2007). Beta-regression coefficients are expressed on the logit scale; for instance, a positive regression coefficient indicates additional increase in the log-odds (and so in the proportion) of the response variable (e.g., DET). Exponentiated coefficients can be interpreted similarly...
to odds-ratios (e.g., an exponentiated coefficient $e^{\beta} = 1.2$ for a deviation-coded binary predictor means that the positively-coded level predicts an odds of DET which is 20% higher than the odds for the negatively-coded level).

The average line length $L$ was modelled using a mixed-effects linear regression. The significance of the predictors was based on non-parametrically bootstrapped (BCa) 95% confidence intervals (10000 iterations) to account for the frequently-attained lower bound on the variable (as for residuals plot inspection). The analysis was implemented in the \texttt{lme4} package (Bates, et al., 2015) in R. Additional details about all model specifications are available in Appendix Q.

To investigate whether the individual-difference measures of PIS, DdC and PfC contributed to lexical convergence between interlocutors, we then conducted a series of Pearson's correlation analyses on the mean and difference z-scores of PIS, DdC and PfC and the CRQA measures of DET, \textit{rENTR} and \textit{L} respectively.

**Results**

**DET**

Contrary to our hypotheses, there was no effect of Partner condition on the extent with which interlocutors lexically converged with each other at word-form level ($\beta = -0.001$, SE = 0.121, $p = .994$, odds-ratio = 1.00, 95% CI [.79; 1.27]). There was, however, a significant main effect of Task ($\beta = -0.49$, SE = 0.12, $p = .0002$, odds-ratio = 1.63, 95% CI [1.27; 2.08]), but no significant Partner-by-Task interaction ($\beta = 0.08$, SE = 0.24, $p = .731$, odds-ratio = 1.09, 95% CI [.68; 1.75]). This indicates that, regardless of Partner condition, interlocutors lexically converged more with each other in the joint-decision-making part compared to the no-goal-oriented part of the discussion task.

**rENTR**

With respect to the predictability in the patterns of converged structures, contrary to our hypotheses there was no effect of Partner condition ($\beta = 0.01$, SE = 0.21, $p = .962$, odds-ratio

---

32 Given the limited degrees of freedom in the data, we did not control for the Individual Difference measures in the mixed-effects regression models.
The main effect of Task was significant ($\beta = -1.39, \text{SE} = 0.24, p < .0001, \text{odds-ratio} = 4.02, 95\% \text{ CI} [2.47; 6.52]$), reflecting higher variability in types of repeated structures in the last 5min task compared to the initial task, contrary to our hypothesis. This held regardless of Partner conditions as indicated by the non-significant Partner-by-Task interaction ($\beta = -0.30, \text{SE} = 0.49, p = .547, \text{odds-ratio} = 1.35, 95\% \text{ CI} [.29; 1.93]$). Note, however, that these results may have been biased by the many missing values in the 5min conversations (see Discussion).

$L$

With respect to the stability of lexical convergence, contrary to our hypotheses there was no effect of Partner condition ($\beta = 0.06, 95\% \text{ BCa CI} [-0.04; 0.20], \text{SE} = 0.06, t = 1.00$). The main effect of Task ($\beta = -0.02, 95\% \text{ BCa CI} [-0.17; 0.06], \text{SE} = 0.06, t = -0.38$) and its interaction with Partner ($\beta = -0.17, 95\% \text{ BCa CI} [-0.47; 0.00], \text{SE} = 0.11, t = -1.49$) were also not significant.

**Figure 23** Observed mean levels of DET, rENTR, and $L$ in the Collectivist (red circle) and Individualistic (blue triangle) condition in the 15min and 5min task. Error bars represents 95% bootstrapped confidence intervals of the observed means. Predicted mean values from the mixed-effects regression models (stars) and horizontally-jittered raw observations are also displayed.
**Lexical Convergence and IDs**

We found no significant correlations between any of the CRQA measures and the dyads' mean or difference scores in PIS, PfC and DdC. Correlation coefficients and Benjamini and Hochberg’s (1995) false-discovery-rate adjusted p-values are reported in Table 13.

**Table 13**

<table>
<thead>
<tr>
<th></th>
<th>PIS</th>
<th></th>
<th>PfC</th>
<th></th>
<th>DdC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean (z)</td>
<td>diff (z)</td>
<td>mean (z)</td>
<td>diff (z)</td>
<td>mean (z)</td>
<td>diff (z)</td>
</tr>
<tr>
<td>DET 15min (n 32)</td>
<td>0.15 (.41)</td>
<td>0.18 (.41)</td>
<td>-0.31 (.41)</td>
<td>-0.22 (.41)</td>
<td>-0.16 (.41)</td>
<td>-0.25 (.41)</td>
</tr>
<tr>
<td>5min (n 32)</td>
<td>0.29 (.16)</td>
<td>0.34 (.12)</td>
<td>0.09 (.61)</td>
<td>-0.35 (.12)</td>
<td>-0.18 (.38)</td>
<td>-0.38 (.12)</td>
</tr>
<tr>
<td>rENTR 15min (n 31)</td>
<td>0.08 (.99)</td>
<td>-0.01 (.99)</td>
<td>-0.02 (.99)</td>
<td>0.13 (.99)</td>
<td>-0.01 (.99)</td>
<td>-0.08 (.99)</td>
</tr>
<tr>
<td>5min (n 14)</td>
<td>-0.27 (.56)</td>
<td>-0.21 (.56)</td>
<td>-0.18 (.56)</td>
<td>0.17 (.56)</td>
<td>0.44 (.34)</td>
<td>-0.59 (.16)</td>
</tr>
<tr>
<td>L 15min (n 32)</td>
<td>0.38 (.19)</td>
<td>0.12 (.85)</td>
<td>0.10 (.85)</td>
<td>0.00 (.99)</td>
<td>-0.14 (.85)</td>
<td>-0.20 (.85)</td>
</tr>
<tr>
<td>5min (n 30)</td>
<td>0.18 (.81)</td>
<td>0.09 (.81)</td>
<td>0.05 (.81)</td>
<td>0.16 (.81)</td>
<td>0.05 (.81)</td>
<td>-0.40 (.81)</td>
</tr>
</tbody>
</table>

**5.2.6.3 Stems lexical convergence**

We then turned to the stems-level transcripts to investigate the more refined form of lexical convergence that is independent of syntactic elements. In our previous analysis at word-form level, in fact, some converged lexical items may not have been captured because of syntactically-dependent morphology (e.g., “did you think [...]?” and “yes, I thought [...]”). Note that removing non-content words yielded a substantial decrease in the number of data points. As a result, we obtained sparse matrices for some of the dialogues in the last 5min task, which prevented us from calculating statistical significance for some of the across-task comparisons on the CRQA measures. For this reason, we decided to analyse lexical convergence on the 20min conversation as a whole.

**Analysis specifications and Results**

We conducted two-sample permutation tests on the CRQA indices of DET, rENTR and L with Partner (Collectivist vs. Individualistic) as independent variable (see Table 14).
Table 14

Descriptive statistics and two-sampled permutation test results comparing the CRQA indices between experimental groups (p-values based on 10,000 monte-carlo samples).

<table>
<thead>
<tr>
<th>Partner Condition</th>
<th>Collectivist M (SD) [n]</th>
<th>Individualistic M (SD) [n]</th>
<th>permuted Welch's t</th>
<th>p-value</th>
<th>Cohen's d [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>DET</td>
<td>8.63 (4.10) [16]</td>
<td>11.46 (3.31) [16]</td>
<td>-2.15</td>
<td>0.039</td>
<td>-0.76 [-1.53; .01]</td>
</tr>
<tr>
<td>rENTR</td>
<td>0.62 (0.15) [9]</td>
<td>0.59 (0.23) [9]</td>
<td>0.38</td>
<td>0.693</td>
<td>0.16 [-2.97; 2.46]</td>
</tr>
<tr>
<td>L</td>
<td>2.09 (0.09) [16]</td>
<td>2.14 (0.18) [16]</td>
<td>-1.02</td>
<td>0.311</td>
<td>-0.35 [-1.11; .39]</td>
</tr>
<tr>
<td>maxL</td>
<td>2.56 (0.51) [16]</td>
<td>3.00 (1.21) [16]</td>
<td>-</td>
<td>-</td>
<td>-0.47 [-2.23; 1.18]</td>
</tr>
<tr>
<td>RR</td>
<td>0.20 (0.04) [16]</td>
<td>0.23 (0.06) [16]</td>
<td>-</td>
<td>-</td>
<td>-0.59 [-2.50; 2.16]</td>
</tr>
</tbody>
</table>

In line with the *linguistic convergence as a compensatory-strategy* hypothesis, interlocutors interacting with a mistrusting individualistic partner showed a higher degree of lexical convergence (DET) than interlocutors interacting with a trusting collectivist individual (p < .05). However, this difference was not convincing when evaluated with Bayesian statistics: BF_10 = 1.85 ± 0% corresponding to anecdotal evidence (Kass & Raftery, 1995). Interlocutors in the two Partner conditions did not show different levels of rENTR or L (all ps > 0.5).

**Lexical Convergence and IDs**

We found no significant correlations between any of the CRQA measures and the dyads’ mean or difference scores in PIS, PfC and DdC. For reasons of conciseness, we are not reporting the analysis results.

**5.2.7 Exploratory analyses**

**5.2.7.1 Windowed CRQA on word-form lexical convergence**

The general CRQA measures reported in the previous section were computed on the entire cross-recurrence plot. In the following analysis, we looked at how lexical convergence might have evolved over time using windowed CRQA, and whether it did so differently in the Individualistic and the Collectivist groups. For instance, previous studies found that pairs of
people who self-reported experiencing higher affinity tended to start their interaction with
greater lexical similarity which then decreased over time; in contrast, low affinity-pairs
started their conversation with less similarity which then increased over time (Liebman &
Gergle, 2016). We decided to restrict this analysis to the word-form transcripts and analyse
the two subtasks separately.

Windowed CRQA depends on the choice of three interconnected parameters: the number of
time points (i.e., how many times we want to sample from the unfolding dialogue), window
size (i.e., how many words should be considered every time we sample) and step (or
sampling interval, i.e., every how many words we sample). The ratio between window size
and step represents the degree of overlap between adjacent windows. As Marwan and
Webber (2015) point out, the choice of these parameters can strongly affect the trend
estimation, and the choice of a suitable value is an acknowledged open issue in the
methodological discussion of (C)RQ analysis (see for instance, Kiselev, 2011).

Given our sample, this posed two challenges. First, the dialogues in our studies differed
substantially in length (word unit) ($M_{15\text{min}} = 515$, range$_{15\text{min}} = [333, 762]$; $M_{5\text{min}} = 166$,
range$_{5\text{min}} = [83, 335]$) and number of turns ($M_{15\text{min}} = 51$, range$_{15\text{min}} = [31, 93]$; $M_{5\text{min}} = 22$,
range$_{5\text{min}} = [10, 70]$), with participants producing on average 10.3 (8.0) words per turn (range
= [6.2, 15.8]; [4.8, 12.6]). To address this issue, we decided to sample all dialogues an equal
number of times while keeping the window-size-to-step ratio constant across dialogues. As a
result, even if the values of window size and sampling interval were unique to each dialogue,
we were able to obtain the same number of normalized time points across dialogues; this
means that, for any time point $n$, $n$ was at a proportionally equal distance (measured in
number of words) $m$ from the start of each dialogue. The window-size-to-conversation-
length ratio was also constant across dialogues.

Second, once we had decided to keep the window size-to-step ratio constant, we had to
determine a value for that ratio. We decide to use the value that overall, across all dialogues
and independently of Partner condition, maximized $RR$. $RR$ can in fact be considered as a
marker of signal (i.e., recurrence)-to-noise ratio. We followed a four-step procedure.
1. We set up an iterative process that, from a starting range of plausible time-points and ratio values, recursively estimated the windowed CRQA measures for each combination of those time-points and ratio for each dialogue. The following requirements had to be met:
   a. \( \text{step} < \text{window size} \) to guarantee overlapping windows. Non-overlapping would have assumed dialogues to be constituted by a sequence of hermetically sealed (linguistic) compartments that do not influence each other.
   b. \( \text{ratio} > 2 \) so that adjacent windows contained at least 50\% of common (linguistic) information.
   c. Given that if the resulting window sizes were too small, they could show spurious information or not cover a sufficient amount of structure in the CRP, and, if too large, they could hide real temporal trends (Kiselev, 2011), we set the \text{window size to have a value between 15-50\% of the length of the dialogue} for the 15min transcripts and \text{between 55\%-80\% for the 5min transcripts}. The lower bounds guaranteed all window sizes contained at least 50 words in all dialogues in both subtasks. Ideally, we would have set for the 15-50\% criteria also for the 5min subtask. However, conversations were too short and so recurrence matrices too small to allow for smaller moving windows and at the same estimate CRQA measures.

Criteria (a)-(c) had to be met by all 32 dialogues.

2. For each dialogue and ratio/timepoints combination, we calculated the geometric mean of the resulting sequence of \( RR \). The geometrical mean is a measure of central tendency commonly used with trend data and time-dependent proportions (Zar, 2010). It follows that for each ratio/timepoints combination, we obtained 32 geometric-mean \( RR \) (one for each dialogue). Figure 24 shows the trend in \( RR \) for a random dialogue/dyad and three random ratio/time-points combinations in each subtask.

3. For each ratio/time-points combination, we took the median of the 32 \( RR \) geometrical means calculated in step 2. Figure 25 shows the density distribution of
the geometrical means for each ratio/time-points combo (i.e., each line corresponds to the density for one combo).

4. We selected the combination of ratio and time-points that was associated with the highest median (i.e., that maximized $RR$ across dialogues).

Figure 26 provides an example that illustrates steps (1)-(3) in the optimization process.
Figure 24 Temporal trend in the %RR in the dialogue of a random dyad for three random ratio/time-points combinations in the 15min (upper panels) and 5min task (bottom panels). Step 2 in the iterative process calculates the geometric mean of %RR in each panel.

Figure 25 Density distribution of the 32 geometrical means for each ratio/time-points combo (i.e., each line corresponds to the density for one combo) in the two subtasks. Vertical lines represent the highest median in each subtask.
The values obtained were ratio = 6 and time-points = 12 for the 15min task (the window-size-to-conversation-length ratio was 0.33), and ratio = 10.3 and time-points = 8 for the 5min task (the window-size-to-conversation-length ratio was 0.56). We used these values to run windowed CRQA on the two sets of transcripts separately.

To explore whether the rate of word-form lexical convergence showed time-dependent patterns, we ran second-order (orthogonal) polynomial mixed-effects beta regressions on the obtained CRQA measures of Determinism (DET). The models also included Partner (Collectivist = .5, Individualistic = -.5) as fixed effects, and by-dyad random intercept and slopes for the first- (t1) and second-order (t2) polynomials to account for the dependencies in the data (see notes in Table 15 for an explanation of the model coefficients). To assess the significance of the predictors in the models, we calculated non-parametrically bootstrapped BCa 95% CI by randomly sampling Dyad (our grouping variable) with replacement, as recommended for longitudinal data (Sherman & le Cessie, 1997). We decided to take the non-parametrical bootstrapped approach given that the growth-curve mixed model showed heteroskedastic errors and the dependent variable had a frequently-attained lower bound (i.e., zero). To compare the effect of the predictors in the two tasks, we computed the t-statistic for the difference between the two models' beta coefficients using Clogg, Petkova and Haritou (1995) formula and taking a |t-statistic| > 2 as significant. Results are reported in Table 15.

---

For the other CRQA measures, there was not enough variability in the data of the 5min dialogues to estimate meaningful model parameters.

The use of the t-over the z-statistic is recommended for small samples (Clogg et al., 1995).

t-value = (b1 – b2) / sqrt(SEb1^2 + SEb2^2), where b1 and b2 are model 1's and model 2's
For both the 15min and the 5min task, the growth-curve mixed model revealed no time-dependent pattern in the extent of lexical convergence (DET) in either Partner condition, although the quadratic term was marginally significant in the open-ended 15min task. Figure 27 illustrates the observed levels of DET over time (averaged across dyads) and the mixed models’ predictions. The effects of the predictors did not differ between the two dialogue tasks (all ts < 2).

**Table 15**

Summary of the two growth-curve mixed-effects beta-regression models for DET\(^{(A)}\), and t-test results for the comparison of the beta coefficients between the models

<table>
<thead>
<tr>
<th>Fixed effects(^{(B)})</th>
<th>Model 15 min task (n = 384, dyad = 32)</th>
<th>Model 5 min task (n = 256, dyad = 32)</th>
<th>Task comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>β</td>
<td>SE</td>
<td>BCa 95% CI</td>
</tr>
<tr>
<td></td>
<td>-2.10</td>
<td>0.07</td>
<td>[-2.19; -1.96]</td>
</tr>
<tr>
<td>t1</td>
<td>-0.25</td>
<td>0.29</td>
<td>[-0.68; 0.23]</td>
</tr>
<tr>
<td>t2</td>
<td>-0.36</td>
<td>0.18</td>
<td>[-0.65; -0.01]</td>
</tr>
<tr>
<td>Condition</td>
<td>0.12</td>
<td>0.15</td>
<td>[-0.11; 0.37]</td>
</tr>
<tr>
<td>t1:Condition</td>
<td>0.11</td>
<td>0.57</td>
<td>[-0.81; 1.05]</td>
</tr>
<tr>
<td>t2:Condition</td>
<td>-0.17</td>
<td>0.37</td>
<td>[-0.79; 0.42]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects (Std Dev)</th>
<th>Intercept</th>
<th>t1</th>
<th>t2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.40</td>
<td>1.56</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>1.06</td>
<td>1.48</td>
<td>0.57</td>
</tr>
</tbody>
</table>

\(^{(A)}\) Before entering the analysis, DET (%) was divided by 100 to obtain proportions. The original range of DET was [0, 0.305] for the 15min task and [0, 0.529] for the 5min task. We applied Smithson and Verkuilen’s (2006) transformation to obtain beta-regression-friendly range: [0.001, 0.306] and [0.004, 0.529] for the two tasks respectively.

\(^{(B)}\) The inverse logit of the intercept captures the average proportion of DET across time-points and Partner conditions. The Partner coefficient represents the difference in DET (on the logit scale) between the two experimental conditions. The first-order polynomial (t1) captures linear trends in the rate of change of DET as dialogues unfold: whether DET increased or decreased compared to the start of the conversation. The second-order polynomial (t2), if significant, indicates that the rate of change of DET changes direction (i.e., the slope changes sign) at a certain point. A positive coefficient for predictor X, and SE the coefficient's standard error.

\(^{36}\) We tried to fit a zero-inflated mixed-effects beta regression using the \{gamlss\} package in R (Rigby & Stasinopoulos, 2005) to account for this challenging combination of elements in our data (i.e., continuous-proportions data, zero-inflation, random slopes). Unfortunately, in simulation work, we found we were not able to use the \texttt{gamlss} function to model the random slopes and obtained inflated Type-I error rates.
coefficient indicates an opened-up curve (i.e., after an initial decrease, DET increases), a negative coefficient indicates an opened-down curve (i.e., after an initial increase, DET drops).

![Graph showing observed rate of change in DET in the two Partner conditions in the 15min (left) and the 5min (right) task. Error bars represent 95% bootstrapped confidence intervals of the observed mean DET at each time point per condition. Lines display the predicted values from the polynomial mixed-effects beta regressions.](image)

**Figure 27** Observed rate of change in DET in the two Partner conditions in the 15min (left) and the 5min (right) task. Error bars represent 95% bootstrapped confidence intervals of the observed mean DET at each time point per condition. Lines display the predicted values from the polynomial mixed-effects beta regressions.

5.2.7.2 Types of lexical convergence

In the analysis at word-form level, we found evidence that interlocutors converged more with each other’s lexical choices when they were engaged in a task with a clear joint decision-making element than a less goal-oriented task (DET results). This was irrespective of Partner condition. However, these results do not tell us whether the greater convergence observed for the 5min joint-decision-making part was simply driven by topicality and task demand (e.g., participants repeating certain words just because they were talking about the same topics) rather than communicative interpersonal convergence (i.e., participants using certain words because their partner had used them; the process driven by social or priming mechanisms). We can assume, in fact, that the goal of agreeing on which objects to take away may have increased the frequency with which participants referred to these objects. To assess this, we compared the score of DET observed for the 32 real dyads with the score of
DET obtained for 32 pseudo-dyads. For each real dyad, we created its pseudo-dyad by keeping one of the speakers and interweaving their dialogue turns with the turns of a speaker from another dyad (cf. Foltz et al., 2015a). This baseline should respect the temporal dynamics and distributional properties of the conversation while disrupting the interpersonal coupling dynamics (see section 4.3.3 for more detail).

We built a mixed-effects beta regression model on the proportion data of DET with Conversation (real = .5 vs. pseudo = -.5), Task (15min = .48 vs. 5min = -.52) and their interaction as fixed-effects and a by-dyad random intercept. Figure 28 shows observed mean value of DET and model-based predictions. We found a small but significant effect of Conversation ($\beta = 0.26, SE = 0.08, p =.003, odds-ratio = 1.30, 95\% CI [1.10; 1.51]$); real interlocutors aligned more than their pseudo counterparts, irrespective of Task demand (Conversation-by-Task interaction: $\beta = -0.06, SE = 0.17, p =.714, ratio of odds-ratio = 0.94, 95\% CI [.68; 1.31]$). The main effect of Task was also significant ($\beta = -.41, SE = 0.08, p < .0001, odds-ratio = 1.51, 95\% CI [1.27; 1.80]$), suggesting that the shorter dialogues in the joint-decision task displayed greater level of lexical repetition irrespective of whether from real or pseudo dyads (see General Discussion).
Figure 28 Observed mean levels of DET in the real (purple triangle) and pseudo (green circle) Conversations in the 15min and 5min task. Error bars represents 95% bootstrapped confidence intervals of the observed means. Predicted mean values from the mixed-effects regression models (stars) and horizontally-jittered raw observations are also displayed.

5.3 GENERAL DISCUSSION

In this study, we aimed to contribute to the understanding of the mechanisms that drive linguistic convergence and tested the effects of social perception on lexical convergence. Our focus was on social accounts according to which linguistic convergence is a response to the impressions we form of the other based on their attitude. Specifically, we tested the claims that linguistic convergence is used either to show and create affiliation and rapport (and anti- or reduced convergence is correspondingly used to show and create disaffiliation and distance) (social distance hypothesis), or as a compensatory strategy to guarantee a smooth interaction with a challenging interlocutor (compensatory strategy hypothesis). We also aimed to replicate previous findings on the role of the interactional goal whereby more goal-oriented conversations lead to greater linguistic convergence. Pairs of participants played a decision-making game with a mistrusting individualistic or a trusting collectivist computer program while believing themselves to be playing with each other. They then joined each
other for a discussion task which was made of a non-goal-oriented subtask and a joint-decision-making subtask. Using CRQA, we compared the extent, stability and predictability of interlocutors' lexical convergence at word-form and stem level according to experimental manipulations.

At no level of analysis did we find reliable evidence that speakers' social perception of their conversational partner, as it was manipulated in the game, influenced interlocutors' overall tendency to lexically converge as measured by either $DET$, $rENTR$ or $L$. Social perception did not affect temporal trends in shorter-term lexical converge either (windowed CRQA results). At the word-form level, however, we observed greater convergence ($DET$) when participants were engaged in the joint-decision-making subtask than the no-goal-oriented subtask, thus replicating previous findings on syntactic convergence (e.g., Reitter et al., 2006). Importantly, when we compared the degree of lexical convergence observed in the real versus the pseudo conversations, we found that real interlocutors showed 30% higher odds to lexically converge, regardless of subtasks. As the pseudo-conversations maintained the temporal and distributional properties of the original but disrupted the coupling dynamics, we assume that this 30% difference is due to interpersonal communicative convergence. That is, participants re-used certain words because their interlocutor had used them and not because they happened to talk about the same things or they would have used them anyway.

Surprisingly, conversations during the joint-decision-making subtask also showed less predictable patterns of lexical convergence ($rENTR$), with participants converging on a larger variety of word combinations. However, we must be wary about these results. Eighteen conversations in the joint-decision-making subtask showed no variability in patterns of repeated words; that is, interlocutors repeated at most two words in a row as displayed by their $L$ and $maxL$ values. For these conversations, $Entropy$ was zero and $rENTR$ could not be estimated. So the comparison of $rENTR$ between subtasks failed to account for the fact that the majority of conversations in the joint-decision-making subtask showed actually highly predictable convergence patterns (i.e., always a maximum of two words in a row). We thus conclude that these findings are potentially misleading and do not discuss them any further.
Finally, we did not find that the individual-difference measures of perceived interpersonal similarity, dominance during conflict and propensity to compromise correlated with any measures of lexical convergence. In the next section, we discuss these results (and lack thereof) in light of previous findings on the influence of social factor on linguistic convergence and the existing theories of linguistic convergence.

5.3.1 On the lack of a social perception effect

In contrast to previous findings on syntactic (e.g., Weatherholz et al., 2014) and phonetic (e.g., Babel, 2012) convergence, we did not find that the attitude displayed by the believed partner in the game (i.e., whether they played in a trusting and collectivist or mistrusting and individualistic manner) affected interlocutors' tendency to converge with each other's choice of words in a subsequent task. These results contrast with both the social distance and the compensatory strategy hypothesis of linguistic convergence, and with social accounts (e.g., CAT) in general. Why was there no effect of social perception?

An easy answer is that our manipulation was not effective and that participants did not form any impressions about the other based on their behaviour in the PGG. However, participants' responses in the post-experiment questionnaire argue against this hypothesis. First, no participant raised any doubt that they might have played the game with a simulated partner. Second, participants described the PGG and what happened during the game emphasising the playing strategy of the other and using terms like “cooperative” versus “selfish” behaviour, “trust”, “making decisions as a team”, “best joint outcome” versus “personal gain”, and “shared success” versus “individual benefits”. Third, when explicitly asked whether the PGG affected their behaviour in the survival communication task, twenty-five people responded that the other's behaviour made them willing to cooperate and feel that they could work well together or the opposite, according to experimental manipulation. Of the thirty-four who did not think they were affected, several specified that this was because the other showed to be cooperative, which confirms that the other's attitude made an impression on them.

Was then our experimental set-up unsuitable to affect linguistic convergence in the predicted ways? There are several reasons why this may be the case. One possibility is that participants might have perceived the same behaviour in different ways, based on their own
interpretation of the game and of the best strategy to keep. For instance, the collectivist behaviour might have been interpreted negatively as a marker of a not particularly strategic person, and not as a positively-valued attempt to maximise shared outcomes. Similarly, the individualistic partner may have come across as smart and strategic rather than mistrusting and selfish, depending on the participant's own understanding of the “best way” to play the game. Within the same dyad, members that may have been differently affected by the manipulation may have pulled the lexical convergence effect in different directions, thus cancelling out the social-perception effect (a possibility also given by our lexical convergence measure which combines the two individual convergence tendencies). As we discussed, previous evidence that (syntactic) convergence is mediated by participants' perception of the other found that the other’s characteristics affected linguistic convergence in relation to the participants’ own attitude (Weatherholtz et al., 2014) or perception of their performance (Lev-Ari, 2015).

Another possibility is that social perception may not affect language convergence beyond the task at hand if the subsequent task is engaging and requires in itself some degree of cooperation. Previous studies that used similar two-stage approaches measured linguistic convergence in tasks that were presented (Lev-Ari, 2015) or perceived (Weatherholtz et al., 2014) as distractor tasks. In contrast, our participants may have been too engaged in the survival scenario and too focused on doing well in the task to be influenced by their previous experience. Moreover, the other's attitude in the new task (e.g., paying attention, reciprocating) may have overcome the impressions participants' formed of “each other” in the PGG. Indeed, previous evidence suggests that interlocutors tend to coordinate language with partners that they are currently finding engaging (Ireland et al., 2011). Note, however, that if so, we should have found a temporal trend in the degree of lexical convergence (e.g., in the individualistic condition, low start followed by an increase) which we did not find. Thus, we are not in a position to give ground to this hypothesis. Nonetheless, the role of engagement on linguistic convergence deserves proper investigation. Looking forward, the choice of a less captivating interactive task may help shedding light on this possibility.

Could there be a role for the medium of communication used? Our participants used a text-based chat with locked turn-taking; they could plan and type their message while their
partner was also typing or posting. Unlike spoken dialogue, where interlocutors can plan their utterances while listening to their partner (e.g., Pickering & Garrod, 2013), our participants may have paid less attention to each other's word selection given that they were busy writing their own messages. In addition, through the act of typing, people might have focused more on what they themselves had carefully planned, adjusted and produced than what the other person had produced. For the reasons just described, this focus on one’s own self production should be stronger, as thus more detrimental to attention, in typed-in chat than in spoken production where comprehension and production are active at once. This lack of attention to each other’s words may have then prevented any effect of social perception on linguistic convergence. Thus, we might have observed different patterns of linguistic convergence if we had our participants verbally interact with each other.

Finally, and in line with what we stressed in Chapter 3, there may be reasons that go beyond pure considerations of experimental paradigm and design. Our results may in fact more simply suggest that converging with the lexical choices of our interlocutor is a tendency that, at least to a certain degree, is unaffected by manipulations of the social context, such as the perception that we have formed of the other in a previous interaction.

We must acknowledge that linguistic convergence may not be an “all or nothing” mechanism that applies uniformly to all lexical choices. For instance, participants in an interactive social-dilemma task showed convergence patterns that selectively depended on the valence of the words and their level of cooperation in the game (i.e., cooperative pairs aligned more on positive words but less on negative words, Scissors et al., 2009; for other selective lexical convergence results see Fusaroli et al., 2012; Nenkova et al., 2008). Our measurements of lexical convergence (i.e., word-form and stems) make no distinction about types of words. It may be, hence, that the influence of social perception on lexical convergence is selectively mediated by the social salience of words: only socially salient words are susceptible to linguistic convergence (and anti-convergence) as an effect of social perception.

Is there any other possible theoretical explanation? Unlike other aspects of speech production (e.g., pitch), lexical choice is a domain where speakers enjoy many degrees of freedom and that rarely represents stable individual preferences (see Lev-Ari, 2015, for a similar discussion on syntax). Words may thus be perceived as less discriminative of the
individual speaker; as such, they may be less susceptible to the influence of social perception regarding alignment than other levels of language (e.g., accent or pitch). Consistent with this claim, previous laboratory studies that found social-perception effects on lexical convergence, albeit of the audience-design type (e.g., Branigan et al., 2011), were based on priming participants with the suitable-yet-dispreferred name alternative for objects. The surprising choice of name may have been identified as peculiar of that specific speaker, thus becoming socially salient and susceptible to linguistic convergence. Indeed, corpora studies on syntactic convergence have observed greater convergence for rare linguistic alternatives (e.g., Reitter et al., 2014). Related to this point, lexical selection cannot simply be reduced to the set of a speaker’s stylistic choices but can heavily affect meaning. So speakers may be less prone to accommodate out of social factors if this affects what they really want to say.

5.3.2 On the findings of an interactional goal effect

A second issue that deserves our attention is the greater lexical convergence observed when participants had to come to an agreement in the discussion task. These results are in line with existing evidence that syntactic convergence is stronger in goal-oriented than more open-ended dialogue (e.g., Reitter et al., 2006; Healey et al., 2014).

One explanation is that the joint-decision-making subtask lent itself to clumps of references to the same items, as speakers tried to agree on which five items to take away with them (cf. Reitter & Moore, 2014, on references to landmarks in Map Task). Also, the shorter 5min conversations led to smaller cross-recurrence matrices in which the short-term convergence component was more dominant; indeed, adjacent utterances have been shown to form clusters of topics which may be reflected in similar word choices across interlocutors. In truth, the real- versus pseudo-dyads comparison indicates that both subtasks led to the same degree of interpersonal communicative convergence. This means that any difference between the two subtasks could be explained in terms of topicality effects and task demands, processes that are not ascribable to coordination dynamics between the interlocutors (e.g., speakers mentioning the name of the found items across several utterances as the task is all about discussing those items). This is further supported by Task being a significant predictor in both real and pseudo conversations.
Could there be any alternative explanation? Having to come to a joint agreement may have led participants to pay more attention to each other's contributions in the last part of the discussion task. They might have read more carefully their partner's arguments and kept in memory what these had suggested, also in light of the fact they could not rely on the chat history. As a result, they might have more deeply processed and encoded each other's word choices, resulting in stronger lexical convergence. In contrast, when they simply debated the survival value of the objects, they could go through the task without remembering or paying too much attention to what the other had said. Indeed, existing evidence shows that people more firmly encode in memory information that they more carefully attended to (Kruschke, 2011) or with a strong emotional component (Monin, 2003). And findings on syntactic convergence have linked stronger convergence tendencies to attention (e.g., Branigan et al., 2007; Weatherholz et al., 2014). At a theoretical level, Reitter et al. (2011) proposed that “engaged” conversations lead to more intense semantic processing; consequently, more lexical material gets held in working memory and is thus more likely to be re-used in subsequent utterances. Similarly, according to the interactive alignment model (Pickering & Garrod, 2004), the greater attention induced by the joint decision-making process would predict stronger alignment in the second subtask because the greater activation which is derived would increase the likelihood of interlocutor re-using the processed lexical material.

Importantly, our results corroborate the view that “environmental” aspects of the conversation, like its communicative function within the task demands, are key in interpreting lexical convergence. Analyses of linguistic convergence should start with clear considerations of the communicative demands of the task at hand as these may affect lower-cognitive mechanisms such as attention and memory encoding, which may be at the basis of the linguistic convergence mechanisms.

5.3.3 On the use(fulness) of CRQA and other methodological challenges

CRQA offers a powerful tool to measure linguistic convergence beyond the artificially constrained paradigms usually employed in dialogue research (e.g., priming of alternatives in picture-description tasks). Moreover, it provides a framework to quantify not only the magnitude of linguistic convergence but also the variability, the temporal evolution and leading-following dynamics in the linguistic coordination processes between two
interlocutors. However, as with most advanced techniques, it presents some challenges. Here, we will briefly discuss those we encountered in the investigation of lexical convergence just presented.

As already highlighted in section 4.2, when applied to behavioural streams of data that present turn-taking and high degrees of freedom in their values (i.e., there are endless combinations of words that speakers can use to express the very same thought), researchers will deal with sparse matrices of cross-recurrence plots – a problem that is exacerbated in short dialogues. Given these sparse matrices, diagonal-based CRQA measures are often impossible to estimate or show little variability that makes statistical comparisons meaningless.

Furthermore, it is important to interpret the individual CRQA measure in the context of the other CRQA measures. As we saw in our study, the comparison between \textit{rENTR} values across conditions could be misleading if not done in the light of the average diagonal length \( L \) scores. Similarly, a high measure of \textit{maxL} could accurately indicate high stability in lexical convergence only if accompanied by high \( L \) and a high number of diagonal lines.

Third, when investigating temporal trends in lexical convergence, windowed CRQA poses the challenge of choosing suitable values for window size, sampling interval and number of time points. This is no trivial issue when dealing with dialogues of various lengths whose sampling and analysis are both at word level. Researchers need a method that allows estimation of the same number of normalised data points across all dialogues. We proposed an iterative approach based on maximising the signal-to-noise ratio across all dialogues and independently of experimental conditions. Looking forward, such an approach should be integrated with a sensitivity analysis.

Finally, we consider effect and sample sizes. Previous findings on the effect of social perception on linguistic convergence at syntactic level report very small effect sizes. Weatherholtz et al. (2014) found an odds-ratio of 1.41 (\( N = 340 \)) for their standard vs non-standard accent manipulation, and Lev-Ari (2015, Experiment 1) reports an effect of 1.35 (\( N = 67 \)) for perceived prestige. Previous lab experiments that looked at the role of audience-design beliefs on lexical convergence found similar ranges of effects, albeit using smaller
samples. For their human vs. computer interlocutor comparison, Branigan et al. (2011) report odds-ratios of 1.60, (N = 16) for the text-based and 1.35 (N = 32) for the spoken interactions. Nonetheless, their priming paradigm presented participants with a choice between two name alternatives, and the constrained interaction only involved processing and producing one word at a time. It is thus highly probable that our sample of 32 dyads was over-optimistic for our unrestricted interactional task, also in view of the between-subjects design.

5.4 CONCLUSIONS

In a two-staged experiment we tested the social accounts’ hypothesis that the degree of linguistic convergence should be affected by the impressions that interlocutors form of each other based on their attitude. At no level of analysis we found reliable evidence that participants’ social perception of their conversational partner (individualistic vs. collectivist), as it was manipulated in the economic decision-making game, influenced their overall tendency to lexically converge as measured by cross-recurrence quantification analysis. At the word-form level, however, we observed greater lexical convergence when participants were engaged in the joint-decision-making subtask than the no-goal-oriented subtask, thus replicating previous findings on syntactic convergence (e.g., Reitter et al., 2006). These results corroborate the view that “environmental” aspects of the conversation, like its communicative function within the task demands, should be afforded a key role when interpreting lexical alignment results (cf. dynamical systems accounts). We did not find any correlation between dyads’ degree of lexical convergence and their members’ individual scores on perceived interpersonal similarity, preference for compromise and dominance during conflict.
Chapter 6

PERSPECTIVE TAKING IN SPATIAL LANGUAGE: A REVIEW

6.0 CHAPTER OVERVIEW

In this chapter, I provide the theoretical background for the experimental work on spatial perspective-taking in language production presented in Chapter 7 and 8. In 6.1 I discuss what is meant by spatial perspective taking in spatial cognition and psycholinguistics, and define the concepts of spatial perspective and spatial reference frames. In 6.2 I present evidence from spatial cognition research that taking a non-self perspective can be cognitively costly and can involve mental simulation. Next, I turn to psycholinguistic research and present findings that people can flexibly juggle between alternative perspectives both when comprehending (6.3) and producing (6.4) spatial descriptions. For production, the evidence suggests explanations related to considerations of communicative efficiency. 6.5 returns to spatial cognition research. I compare findings which link non-self perspective taking to considerations of perceived agency and intentionality against findings that more parsimoniously link this tendency to attention-orienting characteristics of entities that make some entities better sources of spatial coordinates.

6.1 PERSPECTIVE TAKING IN SPACE

Trivially, we all have a perspective on some aspects of the world. I, for instance, know I should improve my Chaturanga Dandasana and I prefer ragù with tagliatelle than with spaghetti. Nonetheless, we often put aside this perspective and recognise that different understandings are possible. Sometimes, we use this information and adjust our communication accordingly. Last time I talked to my mum, I told her I was trying to get better at keeping my body straight parallel to the ground while on my toes and palms and with my elbows at a right angle. And when my British friends asked me what they had to expect for dinner, I happily said tagliatelle with bolognese sauce. We can and sometimes do
take into account others’ informational perspective, in other words what they may or may not know, and what they are able to understand on the basis of their linguistic competence. Similarly, we can and sometimes do account for what a particular object look like from where they stand and whether they can see it (visual perspective), and also for where something is located relative to their position in space (spatial perspective).

In spatial cognition research, spatial perspective taking is defined as the capacity to form a spatial representation of some perceived spatial arrangement. It involves understanding where some elements of the environment are located relative to someone or something else (Kessler & Wang, 2012). Often, it requires resolving a conflict between the position of the self and others within that environment (Surtees, 2011). In psycholinguistics, spatial perspective taking (in language production) has been conceived as the process by which a speaker maps such a spatial representation onto a semantic representation that can be expressed through words (Levelt, 1996). This latter ability is the focus of the current and the next two chapters. As an example, imagine you are watching a medal ceremony at the Olympics game on TV. The three athletes are standing proudly on the podium, facing forward, while the gold medallist’s national anthem is playing. Your three year old child runs into the room, and looking at the TV asks who won. This is an easy question: it is the girl in the middle. Then your child asks who got second. It is the girl on the left from your and your child's point of view, watching the ceremony on TV. It is the girl on the right from the gold medallist’s perspective.

We commonly find ourselves in situations where we must locate objects in a rich perceptual space. Language reflects this richness and allows us to describe space in a variety of ways, explicitly or implicitly locating objects in relation to ourselves, to someone else or to another object (Galati, Michael, Mello, Greenauer, & Avraamides, 2013). There is no one and only way to describe a scene spatially: no one-to-one mapping between the spatial arrangements we perceive and the words we can use to describe them (and the semantic relations that these words underlie) (Levelt, 1996). With this freedom, however, comes also the danger of ambiguity.

For instance, if, looking at Figure 29, I said “I like the teacup on the left”, my spatial description would be ambiguous as two teacups fit the description: one teacup is on my left
as observer but the other teacup is on the Mad Hatter’s left. In order for spatial descriptions to be meaningful, they must rely on a reference frame (RF), i.e., a coordinate system that partitions space (Carlson, 1999), centred on an origin (e.g., ourselves, someone else, an object, or a cardinal point). The origin of the coordinate system determines the system’s vertical (i.e., above/below) and horizontal (i.e., front/behind; left/right) axes and, consequentially, how the corresponding spatial terms (i.e., above, below, front, behind, left, right) must be interpreted.

**Figure 29** The Mad Hatter’s tea party.

Levinson (1996, 2003) distinguishes the relative, the intrinsic and the absolute reference frame (RF) (see also Tenbrink & Kuhn, 2011). Two of these reference frames are relevant

37 Of course I could avoid any ambiguity and identifying the teacup based on a non-ambiguous spatial description (e.g., “the teacup straight in front of the March Hare”) or a non-spatial characteristic “the teacup without the teaspoon”. There is indeed evidence that individuals do find ways to avoid choosing between conflicting spatial perspectives when possible (e.g., Mainwaring, Tversky, Ohgishi, & Schiano, 2003; Tversky & Hard, 2009).

38 Other researchers have defined spatial descriptions based on the nature of the origin of the reference frame. Descriptions anchored in the speaker were called deictic, those centred in unchangeable
for this thesis. In the relative RF, the coordinate system builds in the viewpoint of a perceiver. Spatial relations are defined as ternary relations between this viewpoint, a reference entity (which, crucially, is different from the entity providing the viewpoint) and the located entity. For instance, “The cup behind the teapot”, assuming my vantage point of external observer of the picture. In the intrinsic RF, no viewpoint is involved. An object's location is defined with respect to the directional features of another entity in the environment (reference entity) whose intrinsic axes are used as scaffolding for the coordinate system. Note that this entity can also be you, the speaker. Spatial descriptions are viewpoint-independent and convey binary relations (e.g., “The teacup in front of the March Hare”, given the physical characteristics of its humanised body).

The distinction between a self- and a non-self origin of the coordinate system is independent of the distinction between the two types of reference frame (relative and intrinsic). Importantly, both relative and intrinsic reference frames are compatible with a self origin and with a non-self origin. I can describe the same teacup as the teacup behind the teapot relative to my point of view (self relative), the teacup on the left relative to my own body axis (self intrinsic), the teacup on the right relative to the Mad Hatter’s body axis (non-self intrinsic) or the teacup in front of the teapot from the March Hare’s viewpoint (non-self relative).

In this thesis, we refer to the coordinate system (intrinsic vs. relative) as RFs, and to the origin of the coordinate system as perspective. In our experimental work, we are interested in what may lead speakers to spontaneously abandon their self perspective. The important distinction is thus between self vs. non-self perspectives, which is independent of the distinction between relative and intrinsic reference frames.

As we said, people commonly locate objects in environments that are perceptually rich and that allow a variety of potentially suitable descriptions. Specifically, speakers can choose

________________________
features of the environment as extrinsic, and descriptions centred on any other entity in the environment were called intrinsic (Levitt, 1984; Miller & Johnson-Laird, 1976). We avoid using this terminology given the ambiguity of the term intrinsic, and use self/non-self for descriptions centred on the self/any other entity (we reserve intrinsic for intrinsic RF, following Levinson). Note also that, in the context of interactions, (Schober, 1998) introduced the terms speaker-centred, addressee-centred and object-centred.

39 The non-self origin within the relative RF requires the entity to have a point of view.
among an array of spatial perspectives (the origins of the coordinate systems that will make their spatial terms intelligible). In the experimental work of Chapter 7-8, we investigate some factors that can influence speakers’ choice of perspective: although speakers can always use themselves as the perspective (adopting a default self-to-object representation), they sometimes abandon this perspective in favour of an alternative non-self perspective. Why might they do this?

Insights from the psycholinguistics of dialogue suggest that, under some circumstances (e.g., the presence of a real or imagined addressee), they might do so for communicative reasons (e.g., maximise the resources available for mutual understanding). At the same time, insights from spatial cognition and spatial language processing indicate that perspective taking is a highly adaptive process which is responsive to a diversity of factors, including non-communicative ones. There is, however, no univocal view as to what these non-communicative reasons are. Part of the evidence favours an explanation in terms of perceived agency and intentionality, whereas other findings relate speakers’ choice of non-self perspective to salient features in the environment that make some entities better sources of spatial coordinates. The experiments reported in the next chapters address this duality.

But first, in this chapter, we review the evidence that suggest that taking a non-self perspective can be cognitively costly, especially when it involves processing laterality spatial relations (i.e., whether something is on the left or right of someone/something else). This is particularly relevant to our study as our experimental task involves spatial laterality judgments. We then discuss findings from spatial language processing which suggest that individuals simultaneously activate multiple perspectives when interpreting spatial descriptions. Finally, we discuss findings from interactive and non-interactive research which indicate that also when people talk about (rather than interpret) spatial relations, they show flexibility of perspective and often abandon their own perspective in favour of a non-self perspective, even when this involves computing costly laterality relations.
6.2 **Perspective Taking as a Costly Cognitive Process: Evidence from Spatial Cognition Research**

Research in spatial cognition has long been interested in the mechanisms involved when people adopt a perspective other than the self; how quickly and automatically people process other perspectives when asked to do so (explicit perspective taking) and when not asked to do so (implicit perspective taking); and how these mechanisms are affected or facilitated by characteristics of the environment. In such research, participants have usually been presented with visual scenes that depict a reference entity (usually a human-like figure like an avatar or a doll) and an object to be located. Crucially, the located object is placed in different positions relative to the reference entity so that, on critical trials, the object's location can be judged differently depending on whether one takes one's own perspective or the perspective of the reference entity in the scene.

A technical note to start: a large part of the debate on spatial perspective taking in spatial cognition concerns whether the process is automatic. What makes a process *automatic*? According to the definition generally used in the field, a process is automatic if it does not draw on cognitive resources (or, at least, it poses limited processing demands), it constitutes a direct response to stimuli without being triggered by a conscious decision and needs no additional resources to be initiated and controlled (Bargh, 1994; Surtees, 2011). How to test whether processing a non-self perspective is an automatic process? Studies have so far assumed that if participants show interfering effects (i.e., longer reaction times, more errors or lower acceptability rates) when assessing the appropriateness of a spatial description in contexts where self and non-self perspective conflict (e.g., “the cup on the left”; the same teacup is on my left but on the Mad Hatter’s right) compared to when self and non-self perspectives are not in conflict (e.g., “the teacup in front of the hare”), then these effects are an indication that participants processed both perspectives and that processing a perspective other than one's own occurred outside of cognitive control and unintentionally (as it was not needed or even counter-productive).

Within this line of research, findings suggest that whereas adopting a self (spatial) perspective is automatic and effortless (e.g., Frith & Frith, 2007; Surtees, Apperly, & Samson, 2013; Wang & Spelke, 2002), processing spatial perspectives other than the self is
not a unitary ability. Two types of spatial perspective taking, level 1 and level 2 spatial perspective taking, have been identified based on the different types of cognitive mechanisms that these abilities use. Level-1 spatial perspective taking, which we do not address in this thesis, concerns inferring whether something is located in front of or behind someone/something else. This was observed to rely on an automatic and effortless process (Surtees et al., 2013a), probably facilitated by the ease with which is possible to distinguish an entity’s front/back (when the entity has one) (Bialystok & Codd, 1987; Franklin & Tversky, 1990).

Level-2 spatial perspective taking, on which we focus in this thesis, involves inferring laterality relations and processing whether something is located to the left or to the right of someone/something else. It has been suggested that this second type of non-self spatial perspective taking requires executive control and involves embodied mental self-rotation (probably based on motor representations; Surtees et al., 2013a; Surtees, Apperly, & Samson, 2013b), with individuals imagining themselves assuming the other's position in the environment and simulating the new perspective from there. Evidence for this simulation-based explanation was based on participants assessing whether a sentence cue (e.g., “the ball is to the left of the doll”) or a simpler verbal cue (e.g., ‘left’) matched a given scene while the angular disparity between the participant’s position and the position of the reference entity (i.e., the doll) was manipulated. Participants performed worst (i.e., yielded longer reaction times, more errors) the larger the angular disparity between themselves and the target entity's position. The greater processing difficulties associated with larger disparities were interpreted as a sign that participants were rotating themselves into the doll's position (e.g., Michelon & Zacks, 2006). Furthermore, Kessler and Thomson (2010) showed that the self-rotation is likely to be an embodied process that involves representing one's own body being transposed into the new position and orientation: the processing difficulty increased if there was an initial mismatch between participants’ own and the reference entity’s body orientation (e.g., participants were self-rotated 45° clockwise, whereas the entity was self-rotated 45° anticlockwise; see also Kessler & Rutherford, 2010; Kockler et al., 2010; Zacks

---

40 This distinction closely reflects the one between level-1 visual perspective taking (i.e., processing what someone else can/cannot see) and level-2 visual perspective taking (i.e., processing how something looks like from where someone else stands in space) proposed in developmental psychology (Flavell, Everett, Croft, & Flavell, 1981; Qureshi, Apperly, & Samson, 2010).
Note also that the alternative perspective did not have to be necessarily represented by another person (either as a human body or as an entity with a point of view) (e.g., a chair was sufficient; Michelon & Zacks, 2006; Kessler & Thomson, 2010). In addition, when participants could reply verbally (with “left” and “right”) rather than via key-pressing, processing difficulties were still present but were attenuated (suggesting that when the response was motor-based the effects could be partially explained by stimuli-response incompatibility effects; May & Wendt, 2013).

However, other studies indicate that, under certain circumstances, embodied mental simulation may not be the only strategy used by individuals. Some individuals may employ a “disembodied” route (not necessarily cognitively cheaper; Gronholm, Flynn, Edmonds, & Gardner, 2012; May & Wendt, 2012). For instance, Gardner, Brazier, Edmonds and Gronholm (2013) found that more than half of the participants who assessed whether a black ball was in the left or right hand of a front-view human figure (so that the figure’s left was the participants’ right) solved the task by processing first where the ball was located from their self perspective and then flipping left and right.

In sum, these findings suggest that, at least in scenarios that explicitly require participants to process the spatial perspective of others, individuals may solve laterality judgements (left/right) by putting themselves in the other's shoes (at least when the other’s perspective differ from their own). Nonetheless, individuals seem also to flexibly develop strategies to cope with the task demands, at least under specific circumstances (e.g., when the other perspective represents a 180° rotation from one’s own). Both these lines of evidence are in line with the two-stage accounts of perspective taking proposed within the visual and informational domain. According to these accounts, people stick to their own perspective by default and it is only through an effortful correction of their initial self-based interpretation that they accommodate other perspectives. Importantly, this adjustment would be activated by motivation and controlled by attention (Barr & Keysar, 2002; Epley & Gilovich, 2004; Epley, Keysar, et al., 2004).

Nonetheless, it has also been shown that, with time, practice could lead people to improve on their correcting skills (Epley, Morewedge, & Keysar, 2004). It has indeed been suggested that processes that at the beginning rely heavily on cognitive control (i.e., they are not
automatic), might become automatized with exposure and practice; meaning, their need for attentional cognitive resources and control may reduce (Surtees, 2011). This is an important point. Repeated exposure to and inference about alternative non-self perspectives might make individuals’ ability to adopt a non-self perspective (and the plausible underlying mental simulation mechanisms) only minimally dependent on executive control and attentional resources (Gardner et al., 2013).

6.3 THE SIMULTANEOUS ACTIVATION OF MULTIPLE PERSPECTIVES: EVIDENCE FROM SPATIAL LANGUAGE PROCESSING

Psycholinguistic research in spatial language processing has extensively looked at how individuals interpret spatial descriptions when different perspectives are equally plausible in a given context, and which perspective comprehenders tend to prefer on the basis of different experiential and contextual cues (e.g., Laura Carlson & Van Deman, 2004; Carlson-Radvansky & Logan, 1997; Logan & Sadler, 1996)41. Studies have predominantly focused on the comparison between non-self perspectives that are intrinsic and object-based (e.g., judging something to be in front of a car given the car’s front/back intrinsic axis) and self perspectives that are relative and listener-centred (e.g., judging something to be to the left of the car relative to the listener’s point of view). Studies have also primarily investigated the interpretation of the vertical axis (i.e., the terms above and below), with few exceptions (e.g., Carlson & Van Deman, 2008; Surtees, Noordzij, & Apperly, 2012).

Studies mostly relied on a verbal cue in a visual scene verification paradigm (Clark & Chase, 1972). Participants were exposed to descriptions of the spatial relation between two objects and assessed the suitability of the descriptions with respect to a visual scene either via a binary choice (correct vs. incorrect) or an acceptability scale. Crucially, the scenes depicted a reference entity (an object with directional intrinsic features, e.g., a shoe) relative to which a symmetrical object (e.g., a ball) was to be located. On critical trials, the located object’s position, and so the acceptability of the verbal cue, could be judged differently depending on

---

41 There is an extensive literature in psycholinguistics that looked at the time course and accuracy with which addressees interpret utterances (e.g., “move the candle”) that are ambiguous for them (they see two candles) but not for the speaker (she only sees one of the two candles) (e.g., Barr & Keysar, 2002; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Hanna & Tanenhaus, 2004; Keysar et al., 2000). These studies are not considered in this chapter as they focus on visual and informational perspective taking, rather than spatial perspective taking.
whether participants interpreted the verbal cue from a self (relative) perspective or a non-self (intrinsic) perspective. Under these circumstances, there was no unified evidence whether comprehenders had a clear preference for the non-self (intrinsic) perspective (e.g., Taylor & Rapp, 2004) or the self (relative) perspective (e.g., Carlson-Radvansky & Irwin, 1993; Carlson-Radvansky & Irwin, 1994). Nonetheless, findings point to the importance of contextual factors in the choice of perspective, like the functional relations between the depicted objects (Carlson-Radvansky & Radvansky, 1996; see also Radvansky & Copeland, 2000) and the realism of the scene (Johannsen & De Ruiter, 2013). For instance, Carlson-Radvansky and Radvansky (1996) found that when listeners were presented with pairs of functionally-related objects (e.g., hammer and nail), they tended to prefer non-self spatial descriptions that located one object based on the directional features of the other object (i.e., the non-self intrinsic description “the nail is in front of the hammer” rather than the self relative description “to the right”).

Despite the lack of agreement on which spatial perspective was generally preferred, studies converged in suggesting that comprehenders may simultaneously and automatically activate multiple perspectives when interpreting ambiguous spatial descriptions (Carlson-Radvansky & Irwin, 1993; Carlson-Radvansky & Logan, 1997; Carlson-Radvansky & Irwin, 1994). Specifically, comprehenders were observed to develop alternative understandings of the processed description that enabled them to flexibly respond to changing task demands (e.g., remembering or perceiving the validation scene), and changing spatial configurations (e.g., orientation of the reference entity; Taylor & Rapp, 2004). For instance, in Carlson-Radvansky and Irwin (1993), appropriate self-relative descriptions for scenes that could not be suitably described from a non-self intrinsic perspective were rated as less acceptable compared to descriptions that suitably described the scene from both a self (relative) and a non-self (intrinsic) perspective (i.e., the two perspectives matched). This suggests that, in the former case, the non-self (intrinsic) perspective had been activated and processed by participants even if it did not match the description.

Comprehenders seem then to use cognitive control to select the most appropriate perspective via inhibition (Carlson, West, Taylor, & Herndon, 2002; Carlson-Radvansky & Jiang, 1998). Applying a negative priming paradigm to the spatial verbal cue-picture verification task and
focusing on above/below relations, Carlson-Radvansky and Jiang (1998) observed cognitive costs (longer reaction times) when participants processed a scene from a perspective (e.g., non-self intrinsic) that was plausible but had not been used in the immediately preceding prime trial. This was interpreted as a marker that the perspective had been activated and inhibited in the prime and subsequently re-activated in the probe (Carlson-Radvansky & Jiang, 1998; see also Carlson-Radvansky & van Damen, 2004; Struiksma, Noordzij, & Postma, 2011). Similar mechanisms were found when people processed “front”/”behind” descriptions (Surtees, Noordzij, & Apperly, 2012).

Relevant for the empirical work in this thesis, difficulties in processing laterality relations (left/right) from a non-self intrinsic perspective due to the inhibition/re-activation mechanism were found only when participants had processed left/right relations from a conflicting self (relative) perspective in the preceding prime (Carlson-Radvansky & van Damen, 2008). This indicates that in situations in which the comprehender’s left was the reference entity’s right, both perspectives were automatically activated.

To sum up, theories in spatial language comprehension have argued that when a person is presented with a spatial description, the processed linguistic information is decoded into a mental representation of the spatial arrangements (Taylor & Rapp, 2004; Levinson, 2003). The empirical evidence reviewed in this section suggests that when a person knows she will be asked to compare the linguistic information against a perceptual reality, she will opt for a cognitively flexible strategy and maintain alternative interpretations of this information without endorsing any specific perspective from the start (note that this evidence is in line with the probabilistic-constraints accounts of perspective taking proposed in the interactive language comprehension domain; Brown-Schmidt, 2012; Hanna & Tanenhaus, 2004).

A similar strategy may be used when the same person is presented with a spatial arrangement of objects and is asked to come up with a suitable linguistic description herself. As we mentioned at the beginning of this chapter, “perspective is linguistically free” (Levelt, 1996, p.80). There is no unique way to describe a scene spatially; no one-to-one mapping between what we perceive and the words we can use to describe it. This also means that, most of the time, there is not a single suitable perspective on a scene, but many. Thus, in the next section we turn to findings in spatial language production, and consider whether there is evidence
that speakers adopt cognitively flexible perspective taking strategies when describing spatial relations.

6.4 **Spatial Perspective Taking in Interactive Contexts**

Maybe somewhat surprisingly given the trend in other psycholinguistic fields, experimental work in spatial language production has been biased towards socially interactive contexts; and even when the context has not been per se interactive, speakers have been given an addressee to keep in mind. This bias likely reflects the idea that spatial perspective taking is central to interpersonal communication in that it helps interlocutors construct a common reference framework that allows them to do things together (Beveridge & Pickering, 2013; Hamilton, Kessler, & Creem-Regehr, 2014). Despite the supposed cognitive demand of adopting a non-self perspective (see above), the evidence in spatial language production suggests that interacting with a partner encourage speakers to abandon their self perspective in favour of another’s (usually, the addressee’s) perspective. Speakers engaged in conversation seem to be able to quickly infer where someone else stands within a given environment, and process and describe object locations relative to this someone else, even when doing so involves processing laterality relations.

Very early studies in spatial language production had speakers describe spatial configurations of objects to an imagined addressee (e.g., Bürkle, Nirmaier, & Herrmann, 1986; Herrmann, Bürkle, & Nirmaier, 1987). In these studies, speakers showed a stronger tendency to abandon their self perspective when their addressee’s perspective differed from their own; moreover, they favoured descriptions that could be interpreted unambiguously by both of them, grounded in the intrinsic features of the objects. When speakers were explicitly asked to describe objects’ location from where their addressee (imaginarily) stood, they found it most difficult (manifested in longer speech onset times) when the imagined position was directly opposite them and the object had to be described as on the left/right.

Interactive studies explored how the demands of referential communication tasks (see section 2.5.1) could affect whether speakers described the spatial layout from their own or

---

42 Other studies investigated the perspective that speakers choose to describe routes on a map or through a building (either to an addressee or not) (e.g., Taylor & Tversky, 1992). We won’t consider these studies here as the range of perspectives adopted could not be classified as self vs. non-self.
their partner's perspective. To complete these tasks, pairs of interlocutors holding different vantage points on a set of items (usually symmetrical objects, like circles) helped each other select one of the objects. Importantly, unlike what we saw in spatial language comprehension, the particular paradigms used in language-production research involved conflicting perspectives that belonged to entities (i.e., the two interlocutors) that did not only have bodies with clear directional features but also a point of view.

A classic paradigm is Schober's (1993) object identification task that required participants to verbally localise an item using *left* and *right* in contexts in which left and right had often different interpretations for speaker and listener. Speakers and listeners saw displays of two identical circles, located within a round frame next to each other. Speaker and listener could occupy different positions around the round frame, under relative rotation of 90° or 180°. Their positions were drawn on the same display as the objects. In their descriptions, speakers could use a clear self perspective (“*object to my left*”), a clear non-self perspective (“*object to your right*”), or be ambiguous (“*object on the left*”), leaving the interpretation to the listener.

Findings show that interaction encouraged speakers to adapt their spatial descriptions to the perspective of their partner, with the tendency getting stronger if the partner was only imaginary and, thus, could not negotiate perspective or give feedback on understanding (Mainwaring et al., 2003; Schober, 1993; see also Tenbrink, Maiseyenka, & Moratz, 2007).

In addition, spatial descriptions from a non-self perspective were found to be favoured by speakers’ considerations of what could facilitate the (potential) addressee's task (Levelt, 1984) and by subtle estimations of this latter’s abilities to process and reconstruct the speaker's perspective (Schober, 2009). Schober (2009) showed that speakers with high spatial abilities (with regard to mental rotation) were more likely to provide descriptions from their partner's point of view if this latter was perceived as having low spatial abilities and showed difficulties engaging in the task.

At the same time, a speaker's choice of perspective was also shown to be sensitive to considerations of speaker and addressee’s joint cognitive effort. When asked to remember the spatial layout of a set of symmetrical objects and describe it to a partner that had to
reconstruct it, speakers weighted their and their partner's cognitive demand and flexibly chose the perspective that could minimize collective effort. If their perspectives mismatched only minimally and their partner's perspective was easy to compute (90° offset), they tended to abandon their own perspective in favour of their partner's one. If the mismatch was greater and their partner's perspective harder to compute (135° offset), they stuck to their self perspective (Galati & Avraamides, 2013a, 2013b). Additionally, when there was no mismatch in speaker’s and partner’s perspectives (i.e., self and non-self perspectives would lead to the same spatial description), speakers preferred the self/non-self relative perspective compared to a non-self perspective anchored in the intrinsic axis of a reference object; this tendency was stronger when the partner was actively engaged in conversation (rather than passively receiving the speaker’s monologue-like descriptions; (Johannsen & Ruiter, 2013b).

Speakers’ choice of perspective seems also to be affected by how their conversational partner had previously described spatial arrangements to them. They were observed to abandon their self perspective to a similar degree as their partner had done before for them (Schober, 1995), and to entrain on the same perspective previously used by their partner (Watson, Branigan, & Pickering, 2004). Together, these findings reinforce the idea that speakers handle the “perspective issue” flexibly in conversation and mix between perspectives.

Note that most of these experimental paradigms engaged participants in situations in which the addressee's perspective was essential to the completion of the task. However, pragmatic considerations may play a role in speaker's choice of perspective. In interactive contexts where the partner did not have to act upon the spatial descriptions provided by the speakers, speakers showed weaker tendencies to abandon their self-perspective (e.g., Roche, Dale, & Kreuz, 2010). For instance, in Yoon, Koh, and Brown-Schmidt (2012), participants instructed partners, seated in such a way that left/right had opposite meaning for them, to move objects in a grid between them. When their partner's correct understanding of the description was important to the speaker (i.e., they were requesting him to move the objects), they were more likely to phrase spatial relations according to their partner's perspective (e.g., “to your left”). This was not the case if the speaker's utterance had no real consequences and
they were simply informing their partner about what someone else (the experimenter) would be moving (see also Newman-Norlund et al., 2009).

To summarise, similar to what has been observed for other types of linguistic choices (e.g., referential expressions, e.g., Brennan & Clark, 1996), evidence from interactive (and pseudo-interactive) studies indicates that the way speakers opt for a particular spatial perspective in interaction is a flexible process that is influenced by belief attribution about their addressee (cf. Brown-Schmidt, 2012) and concerns about joint cognitive effort (cf. Clark & Wilkes-Gibbs, 1986), and that is guided by pragmatic considerations. This evidence opens the question of whether this flexible approach to perspective taking in language production is a “mere” by-product of interaction and audience-design considerations. Thus, in the next section, we look more closely at recent findings which suggest that speakers may be induced to spontaneously consider and adopt a perspective other than their own in the absence of any interactive or functional requirements to do so.

6.5 AGENCY, INTENTIONALITY AND ATTENTION-ORIENTING CUES

The evidence reviewed so far suggests that processing perspectives not anchored in the self is cognitively demanding, especially when this involves laterality judgments. For this latter, individuals seem to engage in mental simulation or complex self-to-object transformations. When people describe spatial relations for another person, either in social interaction or in more monologue-like fashions, they take this additional cognitive step and often opt to abandon their self perspective in favour of the other’s. In this section we look at the evidence that suggests that pragmatic and communicative concerns may not be the only explanation for this tendency.

A first set of findings suggests that the presence in the environment of another agent (e.g., a person) who is perceived by the speaker as able to act or currently engaged in action can encourage the coding and the subsequent description of spatial relations from this agent’s perspective.

Tversky and Hard (2009) found that when participants were prompted to describe the position of one object with respect to another one (located on the opposite sides of a table in a photographed scene), they were more likely to abandon their self perspective if the scene
also contained a person who was standing behind the table either looking at or reaching for that object. The tendency was further enhanced if the question asked made explicit reference to an action (e.g., “In relation to the bottle, where is the book placed?”). As the authors suggested, the person's potentiality for action might have induced speakers to abandon their self perspective in an attempt to understand the person' intentions. This was achieved by putting themselves in the other person's shoes and simulating their perspective from there, presumably via embodied mental transformation (Johnson & Demiris, 2005; Thomson & Rutherford, 2010; see section 6.2). Importantly, the increased tendency to abandon the self perspective when the person was present was independent of the task demands (as taking the person's perspective was not functional to what participants had to do) and could not be explained in terms of communicative concerns (as the person was pictured in a scene). Therefore, despite being cognitively demanding, it was argued to be a rather spontaneous process. In line with this agency-based explanation, other studies observed that the likelihood of participants abandoning their self perspective depended on how explicitly the observed person was acting on the object (i.e., gazing vs. grasping) (Mazzarella, Hamilton, Trojano, Mastromauro, & Conson, 2012). For instance, Furlanetto, Cavallo, Manera, Tversky, and Becchio (2013) presented participants with videos of a person who displayed a range from less to more explicit intentions to act on an object (i.e., neutral, gazing, grasping). Participants' tendency to produce non-self based descriptions increased linearly the more explicitly the person was about to act on the object.

However, this agency-based interpretation has not gone unchallenged. Other studies indicate that speakers may adopt a non-self-perspective for reasons that are independent of perceptions of agency and intentionality, and that instead relate to salient physical features of this entity (whether agentive or non-agentive). This evidence largely involves non-linguistic spatial tasks or visual tasks (i.e., concerning what others can perceive rather than where something stands relative to something else). Nonetheless, similar mechanisms might be in place also when speakers verbally locate objects in space.

For instance, Zwickel (2009) showed that participants attributed left/right sides to symmetrical triangles moving on a screen based on the directionality of their motion, and that this attribution interfered with their performance at locating dots on the screen from their
self perspective when, for example, the dot was on their right but the triangle’s left (suggesting that participants computed the triangle's spatial perspective). When presented with pairs of objects in decontextualised contexts, between which there was a functional relationship (e.g., hammer and nail), people tended to describe the location of the one object from a non-self perspective, anchoring descriptions on the other object's intrinsic axes (i.e., “the nail is in front of the hammer”) (Carlson-Radvansky & Radvansky, 1996; see also Coventry & Garrod, 2004). And, in general, speakers tended to describe visual arrangement from a non-self (intrinsic) perspective if one of the objects in the scene had clear intrinsic axes (Johannsen & de Ruiter, 2013a; Ziegler, Johannsen, Swadzba, De Ruiter, & Wachsmuth, 2012); a tendency which was stronger if the non-self perspective allowed front/back descriptions.

Finally, Santiesteban, Catmur, Hopkins, Bird and Heyes (2014) used a modified version of the dot-perspective task, in which individuals verified the number of dots that they could see in a picture while a human-like avatar or an arrow was also in the scene. Participants were slower when the avatar or the arrow was pointing to a subset of all the dots that the participants could see. This indicates that the interfering effects on the mismatched trials were not due to participants simulating the extra entity's (visual) point of view (because the arrow did not have a point of view). Instead these results point towards attention-shifting effects, contingent on the directional features of the reference entity.

In line with this evidence, Heyes (2014) proposed that many entities have features that induce people to code spatial responses according to the entity’s intrinsic axes (e.g., their front/back axis). For example, people’s attention might be drawn to an entity’s front by its eyes or mouth (in the case of a human) or its control panel (in the case of a machine), making it likely that they will use this axis to code a subsequent spatial response. In accord with this account, speakers’ adoption of a non-self perspective would not be related to the perception of agency and intentionality, and it would not occur as a result of simulating how the scene would be perceived by another agent.
To conclude, the tendency of interacting speakers to abandon their self perspective to locate objects can be explained in terms of communicative concerns, especially when the addressee is perceived as less capable of engaging in the task (either because the addressee is not present or because of their abilities). Other findings suggest that this tendency may be underlain by a more “primitive” tendency to take the perspective of agents in general even when speakers are not engaged in interaction or communication with them, in an attempt to make sense of their intentions and action. For both these views, non-self perspective taking is associated with the presence of an intentional entity with a point of view. Making sense of the world from this alternative point of view would be important in view of current (the former) or potential (the latter) interaction.

A third set of evidence, however, indicates the possibility that considerations of agency and intentionality may not always be necessary prerequisites for speakers to abandon their self perspective and point to physical characteristics of the entities in the environment (in line with findings from spatial language processing) that make them a good source of spatial coordinates.

Building on Tversky and Hard’s (2009) paradigm, in Chapter 7 we compare these environmental versus intentional explanations to non-self perspective taking in non-interactive language use. We investigate whether speakers tend to adapt their spatial descriptions to the vantage point of another person out of intentionality-mediated simulation or of general attention-orienting mechanisms.

Building on our results in Chapter 7, in Chapter 8 we take a closer look at the possible role of shared attention on speakers' tendency to abandon their self perspective in spatial language. We build on the findings in joint attention and joint action research that people are highly sensitive to what others are focusing their attention on and are doing, if they have this information. Specifically, evidence suggests that people's own experience is qualitatively affected by the belief that someone else is simultaneously engaged in the same activity (e.g., Böckler, Knoblich, & Sebanz, 2011; Richardson et al., 2012; Shteynberg, 2010), and that when engaged in the same task with someone else, people cannot help but represent the other
person's role in the task (e.g., Atmaca, Sebanz, & Knoblich, 2011; Sebanz, Knoblich, & Prinz, 2005). From this evidence we investigate whether, prior to any communicative and interactional demand, speakers might show a preference for adapting their spatial descriptions to the presumed perspective of someone who (they believe) is attending to the same environment and is engaged in the same task at the same time as them.
Chapter 7

NON-SELF PERSPECTIVE TAKING IN NON-INTERACTIVE LANGUAGE PRODUCTION

7.0 CHAPTER OVERVIEW

This chapter deals with non-self perspective taking in non-interactive language use and compares explanations in terms of intentional vs. purely environmental cues. In three online experiments we investigate whether the presence of another person in the environment leads speakers to abandon their self perspective to describe that space out of intentionality-mediated simulation or general attention-orienting mechanisms. Our results point to the former explanation.

7.1 INTRODUCTION

You are having dinner with friends. There’s a tempting array of dishes scattered across the table. You want to try one of them, but it’s out of your reach, so you ask your friend sitting opposite to pass the dish on the left. People commonly find themselves in such situations, where they must locate objects in a rich perceptual space. Language reflects this richness and allows us to describe space in a variety of ways, explicitly or implicitly locating objects in relation to ourselves (the dish on my right), to someone else (the dish on your left), to another object (the dish in front of the pitcher), or to environmental features (Galati et al., 2013). To do this, speakers must decide on a point of reference or origin that can form the basis of a coordinate system for partitioning space (or reference frame; Carlson, 1999). In this study, we investigate the factors that influence speakers’ choice of origin: although speakers can always use themselves as the origin (adopting a default self-perspective), they sometimes abandon this perspective in favour of an alternative non-self perspective. But what might cause them to do so? Under some circumstances, they might do so for communicative reasons (e.g., to facilitate an addressee’s comprehension), but there are also non-communicative reasons why they might do so.
In this study we consider two types of explanation as to why speakers may spontaneously abandon their self-perspective for non-communicative reasons when producing spatial descriptions. According to the intentional account, they may take another perspective when they detect the alternative perspective of another agent that may have intentions. We contrast this account with the non-intentional account in which speakers’ choice of spatial perspective may be affected only by featural or functional aspects of the environment that are independent of the availability of another agent’s point of view (e.g., a salient environmental feature that attracts speakers’ attention away from themselves and provides a spatial reference point).

To investigate these two possibilities, we conducted three experiments in which participants produced descriptions of the spatial relationship between two objects (e.g., a candle and a pineapple) in a photographed scene (see Figure 1). The objects did not have distinctive front/back or left/right axes, so that participants could not use a perspective based on the objects themselves. We manipulated whether the scene included a non-agentive entity, an agentive entity which was not acting and showed no intention to act, or no additional entity, and also manipulated where that entity was positioned with respect to the two objects. We measured the effect of these manipulations on whether participants used a self-perspective (e.g., “The candle is to the left of the pineapple” to describe a scene such as Figure 30), or whether they abandoned it in favour of a non-self perspective (e.g., “The candle is to the right of the pineapple”).

It is uncontroversial that speakers can adopt a non-self perspective if asked to do so (Bryant, Tversky, & Franklin, 1992), and moreover that they may do so spontaneously in interactive contexts, in which speaker’s behaviour is driven by immediate communicative concerns. For example, in collaborative tasks in which people communicate spatial relationships to each other, speakers are more likely to take a non-self-perspective when they believe that their partner may have difficulty in understanding correctly (e.g., Mainwaring, et al., 2003; Schober, 2009; see also Galati & Avraamides, 2013a), but less likely when the spatial descriptions are not relevant to carrying out the task (Roche, et al., 2010; Yoon, et al., 2012). These effects demonstrate that people do adopt a non-self perspective in the presence of another agent (i.e., the addressee) that has an alternative perspective. However, such (so-
called ‘audience design’) effects do not demonstrate automatic activation of the non-self perspective: They may involve quite explicit reasoning about addressees’ likely understanding and explicit simulation of their self (relative) perspectives, and we do not address them in the current study.

Our concern is rather with findings suggesting that speakers may be influenced by the availability of another agent’s perspective in ways that do not appear to be linked to communicative concerns. That is, there is some evidence that speakers may be triggered to implicitly adopt a non-self perspective simply by the perception of another agent with a different spatial perspective to their own. For example, Tversky and Hard (2009) found that when participants were asked to produce descriptions of objects’ locations in a scene (e.g., in response to the question “In relation to the bottle, where is the book?”) in a non-communicative context, they often abandoned their own perspective if the scene also contained another person with an opposite vantage point on the object who was either looking at or reaching for it.

Further evidence suggests that this tendency may be associated in some ways with speakers’ need to understand other agents’ intentions or actions even when such intentions or actions are not relevant to communication. Tversky and Hard (2009) found that participants’ tendency to adopt a non-self perspective when a person was present in the stimuli was further enhanced if the question made reference to an action (e.g., In relation to the bottle, where is the book placed?). Using similar paradigms, other studies have shown that the likelihood of participants abandoning their self perspective depended on how explicitly the observed person was acting on the object (i.e., gazing vs. grasping) (Mazzarella, et al., 2012). For instance, Furlanetto and colleagues (2013) presented participants with videos of a person who displayed a range from less to more explicit intentions to act on an object (i.e., neutral, gazing, grasping). Participants’ likelihood of producing non-self based descriptions increased linearly as the degree of discernible intentionality increased.

Finally, some evidence suggests that effects of agency sometimes depend on task conditions. In Surtees et al. (2012), participants judged the appropriateness of statements that described the position of a ball with respect to an agent-like (doll) or a non-agent-like (chair) referent. When participants experienced both agent-like and non-agent-like referents in the same
block, they showed an effect of agency: They gave higher ratings for statements that were appropriate under the non-self perspective than to statements that were inappropriate under the non-self perspective (thus displaying sensitivity to the non-self perspective), but importantly this tendency was stronger when the referent was agent-like. However, when participants experienced the two types of referent in separate blocks, the effect of agency disappeared (see also Clements-Stephens, Vasiljevic, Murray, & Shelton, 2013). As these findings demonstrate effects of agency on perspective-taking under some conditions, they provide support for the intentional account.

In summary, some evidence suggests that speakers sometimes implicitly adopt a non-self perspective as a result of perceiving another agent that demonstrates the potential for action, is currently engaged in action, or evinces an intention to act. These results are consistent with the intentional account, and have been interpreted in terms of mental simulation (i.e. a self-centred shift from the observer to the observed person; cf. Johnson & Demiris, 2005).

However, other evidence supports the non-intentional account. Some studies suggest that speakers’ spatial descriptions are influenced by the functional relationships between objects in the absence of any evidence of agency. Carlson-Radvansky and Radvansky (1996) found that when people were presented with pairs of functionally-related objects (e.g., hammer and nail) in a bare context (i.e., with no background or other objects present), they tended to describe the location of the one object using a non-self perspective, based on the directional features of the other object (i.e., the object-centered description “the nail is in front of the hammer” rather than “to the right”; see also Johannsen & De Ruiter, 2013a).

Other evidence from non-linguistic tasks suggests that the properties of inanimate entities can affect spatial perspective-taking. For example, one study showed that participants attributed left/right sides to symmetrical triangles moving on a screen based on the directionality of their motion, and that this attribution interfered with their performance at locating dots on the screen when, for instance, the dot was on their right but the triangle’s left (Zwickel, 2009). More crucially, other findings from non-linguistic tasks challenge the centrality of someone else’s intention to act in spatial perspective taking. Santiesteban and colleagues (2014) had participants count the number of dots on a screen. Sometimes, a human-like avatar or a pointing arrow was present on the screen. Participants were as much
slow and accurate at counting dots when the arrow’s directional features as when the avatar’s directional features (i.e., the arrow’s or avatar’s head) pointed to a subset of the dots, suggesting no modulatory effect of agency.

But how might these low-level attention-orienting mechanisms lead to the use of a non-self perspective in spatial language use? One explanation comes from Heyes (2014), who proposed that many entities have features that induce people to code spatial responses according to the entity’s intrinsic axes (e.g., their front/back axis). For example, people’s attention might be drawn to an entity’s front by its eyes or mouth (in the case of a human) or its control panel (in the case of a machine), making it likely that they will use this axis to code a subsequent spatial response. In accord with the non-intentional account, speakers’ adoption of a non-self perspective does not occur as a result of simulating how the scene would be perceived by another agent.

In sum, there is mixed evidence concerning the factors that may affect speakers’ choice of perspective in spatial descriptions in non-interactive contexts when a person is part of that space. In line with the intentional account of spatial perspective-taking, some recent studies have suggested that when people describe space, the perspective that they adopt is influenced by the presence of another person’s alternative self perspective, as long as this person is perceived as potentially (Tversky & Hard, 2009) or currently (Mazzarella et al., 2012) acting upon the environment in an intentional manner (Furlanetto et al., 2013). However, evidence from other studies (largely involving non-linguistic spatial tasks) supports the non-intentional account by suggesting that speakers may adopt a non-self-perspective for reasons that are independent of perceptions of agency and intentionality.

7.1.1 The current study

Our study therefore investigates what may cause people to switch perspective in non-interactive spatial language. Specifically, what characteristics of an entity affect whether speakers describe the space within which it appears using a non-self perspective rather than a self perspective? Are speakers influenced by the actual or potential agency of an entity? Or are speakers influenced instead by characteristics of an entity that make it a good source of spatial coordinates (cf. Heyes, 2014)?
Building on the task used by Tversky and Hard (2009), we had participants describe the spatial relation between two objects located on a table in response to the question *On which side of the X is the Y?*. The objects to be described had no inherent or functional asymmetries (i.e., they were symmetrical), and no functional relationship to each other. Sometimes, the scene also contained an agentive entity (i.e., a person) or a non-agentive symmetrical entity (i.e., a plant). Even if inherently spatial, our task did not ask participants to take account of the spatial relation between the objects and the additional entity (the questions always and only referred to the two objects on the table). This is important as experimental settings that only displayed one object on top of the additional key entity (e.g., a person) may have induced participants to base their spatial descriptions on this extra entity, making the task inadvertently biased towards a non-self perspective (e.g., Mazzarella et al., 2012).

In Experiment 5, we tested whether humans, in a neutral pose and with an orientation that did not correspond to the participant’s orientation with respect to the scene (agentive entity condition), elicited non-self spatial perspective taking to a greater extent than plants (non-agentive entity condition) and when no additional entity was present in the scene (no entity condition), and whether any difference was modulated by whether the human could see and potentially act on the object. In Experiment 6, we manipulated the order of presentation of the three entity conditions (agentive vs. non-agentive vs. no entity) using a blocked design, in order to test whether such effects were influenced by the experiential context. Specifically, we examined whether participants used a non-self perspective in non-agentive trials only when they had used that perspective before in agentive trials. In Experiment 7, we manipulated the orientation of the human so that it faced in the same versus the opposite direction to the participants, in order to examine whether another agent’s perspective also affected a speaker’s choice of spatial description when this perspective corresponded to the speaker’s self perspective (i.e., did not constitute an alternative perspective).

### 7.2 Experiment 5

In this experiment, participants produced spatial descriptions for scenes containing an agentive entity, a non-agentive entity, or no additional entity (Entity manipulation). The agentive entity was a person in a stationary pose. The person was facing forward and their arms were crossed; they were not engaged in action and evinced no clear intention to act
(e.g., reaching for or looking at an object). This pose also ensured that the person displayed no attention-orienting cues associated with body asymmetries caused by specific postures (e.g., head turned, one arm stretched out). The non-agentive entity was a plant, which was closely matched in size to the person and was symmetrical. We assume that people do not regard this plant (or, most likely, plants in general) as intentional or as having a point of view or a distinct front/back or left/right sides. We also manipulated whether the agentive or non-agentive entity appeared in front of or behind the objects (Location manipulation). Crucially, the agentive entity always faced the participant so that its orientation and viewpoint were the opposite of the participant’s (see Figure 30). As a consequence, the entity could potentially see and act on the objects only in the agent-behind condition (Figure 30d).

According to one version of the intentional account (perceived intentional account), participants should tend to produce a non-self description (e.g., “the candle is to the right of the pineapple”) when they perceive the entity to be intentional and when this entity would take a perspective on the objects that would correspond to a non-self perspective, and this tendency does not depend on an observable action or intention to act. This should be the case in the agent-behind condition (Figure 30d). It would not be the case in the agent-front condition (Figure 30c) because the entity would not be able to see or act on the objects. More obviously, it should not occur in either of the non-agent conditions; in the non-agent-behind condition, even if participants were to take the plant’s perspective, it is not clear why they would assume the plant to be facing them.

According to another version of the intentional account (actional intentional account), the simple availability of an agent’s alternative perspective is not sufficient to trigger the assumption of intentionality and thus a non-self perspective; instead, an action or a clear intention to act are needed to trigger that assumption and so activate the non-self spatial perspective in a language task. If so, then participants should be equally likely to produce a non-self description in the agentive conditions (Figure 30c-d), the non-agentive conditions (Figure 30a-b), and the no-entity condition (Figure 30e), because the person and the plant in our scenes display no action or intent to act (i.e., no effect of entity or entity-by-location interaction). Importantly, either pattern of results would implicate the involvement of some form of agency-dependent simulation mechanism. In the perceived intentional account,
simulation is triggered by an entity’s action potential; in the actional intentional account, it is triggered by evidence of an intention to act or engagement in action.

Under the non-intentional account, in contrast, both agentive conditions should trigger more non-self perspectives than the plant conditions and the no entity condition (as plants have no salient intrinsic vertical axes); in addition, there should be no difference between the agent-behind and the agent-front conditions because the directional features of the human body do not change between the two.

7.2.1 Method

7.2.1.1 Participants

We recruited 144 participants on CrowdFlower, a crowdsourcing self-service platform that allows clients to design and execute their own workflow (and allows full anonymity). All participants were English native speakers and were paid $0.40.

7.2.1.2 Materials

We used twelve experimental items. Each item was constituted by a pair of experimental objects that had no clear front/behind or left/right (e.g., candle/pineapple). We constructed two sets of five photographed scenes for each item: one set for each possible position of the two objects on the table (i.e., candle on the right and pineapple on the left vs. candle on the left and pineapple on the right); five photographs, corresponding to the four combinations of Entity (Agent vs. Non-agent) and Location (Front vs. Behind), and the No Entity control. See Figure 30. In the Entity conditions, we used four plants and four people (two female, two male). Each scene was paired with a question of the form “On which side of X is Y?” (e.g., “On which side of the candle is the pineapple?”). Across the item set, each object appeared on the left and right an equal number of times, and was asked about an equal number of times; each person/plant appeared front/behind an equal number of times.

We also constructed 16 filler items, each displaying two further objects on a table. In half the fillers, objects were different colours (e.g., a red baking tin and a green salad bowl) and participants were asked “What colour is the X?” (e.g., “What colour is the salad bowl?”). In the other half, objects were of different sizes (e.g., a tube of toothpaste and a sewing machine)
and participants were asked “With respect to the X, how big is the Y?” (e.g., “With respect to the toothpaste, how big is the sewing machine?”).

We created 12 lists, each comprising 24 experimental and 16 filler trials. Each list comprised eight Entity /Agent trials (four Front, four Behind), 8 Entity /Non-agent trials (four Front, four Behind) and eight No Entity trials. In each list, each of the twelve experimental items was assigned to two of the conditions, in such a way that it never appeared twice in the same Entity/Location condition. The sequencing of experimental and filler trials was fixed across all lists so that there were never more than two consecutive experimental trials. We created six versions of each list by randomizing the order of experimental items and filler items, with the constraint that across lists, the first trial was an Agentive, Non-agentive, or No Entity trial an equal number of times (i.e., six), yielding 72 lists. Participants were randomly assigned to one of the lists (two participants per list).

Figure 30 Example of a set of photographed scenes (item candle/pineapple) used in Experiment 5 and in Experiment 6: (a) Non-agent Front; (b) Non-agent Behind; (c) Agent Front; (d) Agent Behind; (e) No Entity.

7.2.1.3 Design

We manipulated: whether an Agent (person) or a Non-agent (plant) was present in the scene between the two objects (additional Entity); whether the entity was located in front or behind
the objects (Location). The study used an incomplete within-subjects factorial design with No Entity as baseline and a 2×2 design with Entity (Agent vs. Non-agent) and Location (Front vs. Behind) as factors.

7.2.1.4 Procedure

On the CrowdFlower platform, participants read the information introducing the study and gave their informed consent to take part. They were then asked to follow a link to an external website, hosted on the university's server, where the experiment was run. By clicking on the link, they were randomly assigned to one of the 72 experimental lists. The first page described the task, informing participants that they would see a series of photographs of scenes and would be asked a question about their content. Scenes were 600×400px and were displayed in the centre of the screen. The question (e.g., “On which side of the candle is the pineapple?”) appeared on the screen below the scene at the same time as the scene itself. Participants typed in their answer and pressed Enter to continue on to the next trial.

7.2.2 Analysis

7.2.2.1 Coding and exclusion criteria

Twelve participants provided no responses. We eliminated 72 of the experimental trials because the photo did not upload, participants typed in nonsense strings of characters, or participants did not answer the question asked. We ended up with 3096 data points from 132 subjects (1035 in the No Entity, 1017 in the Agent, and 1043 in the No-Agent condition). Of the 3096 responses, 3095 were categorized as mentioning that the object was on the right (e.g., “right”, “right side”, “to the right”) or on the left (e.g., “left”, “left side”, “to the left”) with respect to the other object on the table. One response mentioned both left and right (i.e., “my left”, “his right”) and was excluded. Responses were coded as ‘self’ if the position of the object was described from the participant's perspective and ‘non-self’ otherwise. As we were interested in the factors that underlie people’s decision to use a non-self perspective in spatial descriptions, we excluded from the analysis all those participants (44, corresponding to 30.5% of data) who never produced both perspectives at least once. The data for analysis contained 87 subjects for a total of 2131 observations (691 in the No Entity, 675 in the Person, and 693 in the Plant condition).
7.2.2.2 Models

We used mixed logit regression analysis with crossed random effects (Jaeger, 2008) to predict participants’ likelihood to describe the spatial location of objects from a non-self perspective (DV: ‘non-self’ response = 1; ‘self’ response = 0) based on Entity and Location. The crossed hierarchical nature of the model allows explicit estimation of by-participant and by-item individual differences. Predictors were coded with deviation contrasts (levels: -.5/.5) to reduce collinearity and obtain ANOVA-like interpretation of main effects. Their regression coefficient represents the difference on the log-odds scale in producing a non-self description between the two levels of the predictor. We also report odds-ratio (i.e., \( \exp^{\text{log-odds}} \)).

The maximal by-participants and by-items random structure justified by the design was included to avoid anti-conservatism (Barr et al., 2013). All models were fitted using the lme4 package (version 1.1.11, Bates et al., 2015) in R (version 3.2.1, R Code Team, 2014). In the event of failed convergence or perfect correlation (\( r = -1/1 \)) between all the by-participants or by-item random effects, a sign of over-parameterization, we re-fitted the model after forcing independence between those effects. The significance of the predictors (i.e., the model fixed-effect parameter estimates) was evaluated via Wald's Z statistic, which assesses whether the fixed-effects coefficients are significantly different from zero (given the estimated S.E.).

7.2.3 Results

Overall, participants who used both self and non-self perspectives during the experiment took a non-self perspective on 15.6% of trials. Table 16 and Figure 31 show frequency and proportion of non-self descriptions by Entity and Location condition.
Table 16
Frequency of non-self and self perspectives by Entity and Location (N = 1967) in Experiment 5

<table>
<thead>
<tr>
<th>Entity</th>
<th>Location</th>
<th>Self</th>
<th>Non-self</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>Front</td>
<td>279</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Behind</td>
<td>277</td>
<td>70</td>
</tr>
<tr>
<td>Non-agent</td>
<td>Front</td>
<td>302</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Behind</td>
<td>285</td>
<td>61</td>
</tr>
<tr>
<td>No Entity</td>
<td></td>
<td>594</td>
<td>97</td>
</tr>
</tbody>
</table>

Figure 31 Observed proportions of non-self descriptions in the five experimental conditions (Experiment 5). Error bars represent non-parametrically bootstrapped (BCa) 95% confidence intervals on the participant-wise conditional means.

In order to assess whether the presence in the scene of an agentive vs. non-agentive additional entity influenced participants’ perspective, and whether this was further mediated by the location of the entity with respect to the objects, we ran a mixed-effects logistic
regression with Entity (Agent = .5/Non-agent = -.5), Location (Behind = .5/ Front = -.5) and their interaction as fixed effects, and maximal random structure. The results showed that participants were more likely to take a non-self perspective when an additional entity appeared in the scene, located behind the table, as shown by the significant positive coefficient for Location (p = .01). Location did not interact with Entity (p > .9), indicating that the effect was not mediated by whether or not the additional entity was agentive (p > .8). The summary of the model is given in Table 17.

To confirm that our data favoured the null hypothesis of no interaction between Entity and Location, we performed a Bayes Factor analysis, which quantifies the likelihood of observing the data if agentive and non-agentive entities lead to different patterns of non-self descriptions in the behind and front conditions, compared to if they do not. We constructed the null-model, a GLMM with only the main fixed effects of Entity and Location; this model assumes that agents and non-agents do not differently affect perspective in the behind and front locations. The alternative model is our original model that includes the Entity by Location interaction. We then used the two models’ Bayesian Information Criterion (BIC) values to estimate the Bayes Factor as $e^{(\text{BIC}_{\text{alternative}} - \text{BIC}_{\text{null}})/2}$ (Wagenmakers, 2007; Masson, 2011). The null model (i.e., without the Location by Entity interaction) fit the data better by a Bayes Factor of $e^{(1192.0 - 1184.7)/2} = 38.5$, providing strong evidence against the hypothesis that the agentive entity led to a different patterns of spatial descriptions in the behind versus front location compared to the non-agentive entity (posterior probability in favour of the null-model BF/(BF + 1) = .97, which represents strong evidence according to Kass and Raftery’s, (1995) categorization.
Table 17
Summary of the mixed logit model with Entity (Agent = .5 / Non-agent = -.5) and Location (Behind = .5 / Front = -.5) as predictors for Experiment 5 (N=1368)

<table>
<thead>
<tr>
<th>Predictors (fixed effects)</th>
<th>Parameter estimates</th>
<th>Wald's test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log-odds ($\beta$)</td>
<td>S.E.</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.44</td>
<td>.23</td>
</tr>
<tr>
<td>Entity</td>
<td>0.07</td>
<td>.37</td>
</tr>
<tr>
<td>Location</td>
<td>1.02</td>
<td>.42</td>
</tr>
<tr>
<td>Entity: Location</td>
<td>0.01</td>
<td>.76</td>
</tr>
</tbody>
</table>

Random effects Explained standard deviation estimates

| Subjects:(intercept)       | 1.44                |
| Subjects: Entity           | 1.42 .21            |
| Subject: Position          | 0.69 -.96 -.27      |
| Subject: Ent:Loc           | 1.50 -.12 -.76 .35  |
| Items:(intercept)          | 0.19                |
| Items: Entity              | 0.26 0.78           |
| Items: Location            | 0.83 1.00 0.75      |
| Items: Ent:Loc             | 1.03 -0.17 -0.75 -0.13 |

1 All correlations between fixed-effects $< |.59|

To verify whether participants produced more non-self perspectives when any additional entity was present in the scene than when no additional entity was present (cf. Tversky & Hard, 2009), we performed an additional mixed logit analysis comparing the No Entity baseline with the individual four levels of the 2×2 design. To do this, we created one factor representing each combination of Entity and Location, and then converted this factor to a deviation-coding numeric representation. The model summary is presented in Table 18. No difference was observed for either type of entity in the Front condition (ps > .5), but both types triggered significantly more non-self perspective responses in the Behind condition compared to the No Entity baseline.
Table 18
Summary of the mixed logit model comparing the No Entity baseline and the four unique combinations of Entity and Location via deviation coding in Experiment 5 (N=2058)\textsuperscript{1}. Regression coefficients represent the difference to produce a non-self description (on the logit scale) between the specified Entity-Location and the No Entity.

<table>
<thead>
<tr>
<th>Predictors (fixed effects)</th>
<th>Parameter estimates</th>
<th>Wald's test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log-odds ($\beta$)</td>
<td>S.E.</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.45</td>
<td>0.21</td>
</tr>
<tr>
<td>Agent/Front (d)</td>
<td>-0.14</td>
<td>0.54</td>
</tr>
<tr>
<td>Agent/Behind (d)</td>
<td>0.91</td>
<td>0.39</td>
</tr>
<tr>
<td>Non-agent/Front (d)</td>
<td>-0.18</td>
<td>0.50</td>
</tr>
<tr>
<td>Non-agent/Behind (d)</td>
<td>0.81</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Random effects | Explained standard deviation estimates

| Subjects: (intercept)     | 1.44 |
| Subjects: Ag/Fro          | 1.66 | -0.01 |
| Subject: Ag/Beh           | 1.15 | -0.56 | 0.71 |
| Subject: NnAg/Fro         | 1.18 | -0.26 | -0.03 | 0.23 |
| Subject: NnAg/Fro         | 1.15 | -0.72 | 0.20 | 0.54 | 0.79 |
| Items: (intercept)        | 0.19 |
| Items: Ag/Fro             | 0.42 |
| Items: Ag/Beh             | 0.61 |
| Items: NnAg/Fro           | 0.51 |
| Items: NnAg/Beh           | 0.46 |

\textsuperscript{1} Because of failed convergence of the model with maximal random structure justified by the design, we refitted the model by forcing independence between the by-item random effects. Highest correlation between fixed effects, $r = 0.49$.

7.2.4 Discussion

When describing the location of an object relative to another object, participants were more likely to adopt a non-self perspective when the scene contained an additional entity than when it did not. This tendency occurred when the additional entity appeared behind the table (i.e., on the other side of the table relative to the participant), but not when it appeared in
front of the table (on the same side of the table relative to the participant). However, this tendency was not affected by whether the additional entity was agentive (a human) or non-agentive (a plant).

The pattern of results is not compatible with the non-intentional account, under which any entity (agentive or non-agentive) may be spontaneously used as a source of spatial encoding, provided that it has directional intrinsic features that could attract attention and be used as the origin of a coordinate system. Our plants did not have such features (and so could not be used as the origin of a coordinate system), and the directional features of the human body were the same in the agent-front and agent-behind conditions (and so are equally available as the origin in both conditions).

The fact that the non-self perspective effect was limited in the agent conditions to the behind position provides some support for the intentional account. Specifically, it supports the perceived intentional account, in which the presence of an agent can trigger the adoption of a non-self perspective. It does not however support the actional intentional account, in which the presence of an agent can trigger the adoption of a non-self perspective only when it evinces a current action or intention to act. However, the presence of the effect in the non-agent-behind condition is only compatible with the perceived intentional account if participants treated the non-agent as intentional and forward-facing.

One possibility is that participants did interpret the non-agent in this way, and that they did so because they transferred its intentionality and direction from the agentive condition. Some research suggests that spatial perspective-taking can be affected by experimental experience. Clements-Stephens et al. (2013) found that the orientation of featureless triangles affected participants’ approach to a visual perspective-taking task (i.e., recognizing from the position of which triangle a series of pictures had been taken), but only when they had previously been exposed to triangles depicted with eyes. In a similar way, participants might have adopted a non-self perspective for trials containing a person, and then transferred this perspective to scenes that included plants appearing in the same location as the person. To test this possibility, we controlled for possible contamination effects between entity conditions and conducted Experiment 6 in which we presented the agentive (person) stimuli and the non-agentive (plant) stimuli in blocks.
7.3 EXPERIMENT 6

Experiment 6 manipulated the same factors as Experiment 5, but the three Entity conditions were presented in different blocks, with blocks counterbalanced for order of presentation across participants. If people produced more non-self perspective descriptions for the Non-agent-Behind stimuli in Experiment 5 (compared to the Non-agent-Front condition and the No Entity condition) because they transferred intentionality/orientation from the Agent stimuli, then we should find no such tendency in Experiment 6 when we controlled for possible contamination effects. However, we would still expect to find a tendency for participants to produce more non-self descriptions for the agent-behind stimuli compared to the Agent-Front and the No Entity condition.

7.3.1 Methods

The experiment was identical to Experiment 5 except for the following details.

7.3.1.1 Participants

We recruited 288 participants from the same population as Experiment 5 to take part in the online experiment, under the same conditions.

7.3.1.2 Materials

We used the same stimuli and the same ratio between experimental and filler trials as in Experiment 5, with 8 experimental and 5 filler trials per block. We constructed six versions of each of the 12 lists of trials, one version for each order of the blocks (Nonagent-Agent-NoEntity; Nonagent-NoEntity-Agent; Agent-Nonagent-NoEntity; Agent-NoEntity-Nonagent; NoEntity-Nonagent-Agent; NoEntity-Agent-Nonagent), for a total of 72 lists. Four participants were randomly assigned to each list. To avoid carryover effects, a filler trial was always placed between the end of a block and the beginning of the next one. We also created three sequences of experimental and filler trials making sure that there were no more than two consecutive experimental trials, and that each block order was combined with each sequence order.
7.3.1.3  Procedure

At the end of the experiment, participants reported their gender, first language, and age.

7.3.2  Analysis

7.3.2.1  Coding and exclusion criteria

We excluded 52 participants (18%) who had taken part in Experiment 5, 23 participants (8%) for whom English was not their first language, and three participants (1%) who provided no responses. We further excluded one participant who acknowledged that more than one perspective was possible but did not provide any (e.g., “It all depends on perspective”), one participant who described the location of the objects without taking any perspective (e.g., “opposite”), and one participant who explicitly took both perspectives (e.g., “my right, the cup’s left”) for all experimental trials. As in Experiment 5, we excluded from analysis all those participants (92, corresponding to 31.4% of data) who never produced both perspectives at least once. We fitted a generalized linear model (GLM) with logit link function using the glm function in R. The six-level BlockOrder predictor was sum-contrast coded so that the intercept represented the mean of the logit means (i.e., a measure of the average likelihood of never producing a non-self perspective across all block orders). The regression coefficients represented the difference on the logit scale between the corresponding block order and that average. When a categorical predictor is sum-contrasts coded, one of its levels is not compared against the grand mean (the intercept). We chose the block order whose estimated logit (descriptively) differed the least from the mean of the logit mean (i.e., A-N-Na). The GLM results showed that the NoEntity-Nonagent-Agent block order significantly increased the likelihood of participants never switching to a non-self perspective.

\[\text{Log-odds (N-S)} = \begin{matrix}
\text{Switchers} & 18 & 23 & 14 & 23 & 23 & 18 \\
\text{Non-switchers} & 14 & 12 & 22 & 12 & 15 & 13 \\
\end{matrix} \]

\[\text{Mean of logit means} = \begin{matrix}
\text{Switchers} & .25 & -.65 & .45 & -.65 & -.43 & -.33 & -.31 \\
\end{matrix} \]
technical issues with photo uploading. Data from the 115 participants who produced both a self and a non-self description at least once were analysed, yielding a total of 2757 observations (920 in No Referent, 919 in Agent and 918 in Non-agent conditions).

7.3.3 Results

On average, participants took a non-self perspective on 17% of trials. Table 19 and Figure 32 show the frequency and proportion of non-self descriptions by Entity and Location condition.

Table 19

Frequency of non-self and self descriptions by Entity and Location (N = 2757) in Experiment 6

<table>
<thead>
<tr>
<th>Entity</th>
<th>Location</th>
<th>Self</th>
<th>Non-self</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>Front</td>
<td>375</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Behind</td>
<td>345</td>
<td>112</td>
</tr>
<tr>
<td>Non-agent</td>
<td>Front</td>
<td>398</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Behind</td>
<td>397</td>
<td>61</td>
</tr>
<tr>
<td>No Entity</td>
<td></td>
<td>782</td>
<td>137</td>
</tr>
</tbody>
</table>

compared to the average likelihood across all orders ($\beta = .76$, SE = .31, $p = .015$, odds-ratio = 2.1). All other comparisons were not significant ($ps > .2$).
A logit mixed model analysis was used to model outcomes. No effect of Entity (p > .3) or Location (p > .1) was found. Crucially, there was a significant interaction between these two factors (p < .05); participants were more likely to take a non-self perspective to describe the spatial location of objects when an agentive entity but not a non-agentive entity was in the scene, and was located behind the objects. A summary of the model is given in Table 20.
Table 20
Summary of the mixed logit model with Entity (Agent = .5, Non-agent = -.5) and Location (Behind = .5, Front = -.5) as predictors for Experiment 6 (N=1835)

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Parameter estimates</th>
<th>Wald's test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log-odds (β)</td>
<td>S.E.</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.46</td>
<td>0.20</td>
</tr>
<tr>
<td>Entity</td>
<td>0.43</td>
<td>0.39</td>
</tr>
<tr>
<td>Location</td>
<td>0.36</td>
<td>0.33</td>
</tr>
<tr>
<td>Entity:Location</td>
<td>1.53</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Random effects: Explained standard deviation estimates

| Subjects: (intercept) | 1.52 |
| Subjects: Entity     | 1.81  | 0.30 |
| Subject: Location    | 1.02  | -0.22 | -0.24 |
| Subject: Ent:Loc     | 1.38  | -0.97 | -0.27  | 0.44 |
| Items: (intercept)   | 0.00  |
| Items: Entity        | 0.64  |
| Items: Location      | 0.33  |
| Items: Ent:Loc       | 0.86  |

1 Highest correlation between fixed effects < |.50|.

As with Experiment 5, we then tested whether participants’ likelihood of describing the scenes from a non-self perspective differed in the four entity conditions compared to when no additional entity was in the scene. We conducted a mixed logit analysis comparing the No Entity baseline with the individual four levels of the 2×2 design (see section 7.2.3 for details). The model summary is presented in Table 21. No difference was observed for either conditions (ps > .1).
Table 21

Summary of the mixed logit model comparing the No Entity baseline and the four unique combinations of Entity and Location via deviation coding in Experiment 6 (N=2754)\(^1\). Regression coefficients represent the difference to produce a non-self description (on the logit scale) between the specified Entity-Location and the No Entity.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Parameter estimates</th>
<th>Wald's test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log-odds ((\beta))</td>
<td>S.E.</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.39</td>
<td>0.18</td>
</tr>
<tr>
<td>Agent/Front (d)</td>
<td>-0.39</td>
<td>0.45</td>
</tr>
<tr>
<td>Agent/Behind (d)</td>
<td>0.56</td>
<td>0.36</td>
</tr>
<tr>
<td>Non-agent/Front (d)</td>
<td>-0.07</td>
<td>0.31</td>
</tr>
<tr>
<td>Non-agent/Behind (d)</td>
<td>-0.35</td>
<td>0.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Explained standard deviation estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects: (intercept)</td>
<td>1.39</td>
</tr>
<tr>
<td>Subjects: Ag/Fro</td>
<td>2.31</td>
</tr>
<tr>
<td>Subject: Ag/Beh</td>
<td>1.68</td>
</tr>
<tr>
<td>Subject: NnAg/Fro</td>
<td>0.82</td>
</tr>
<tr>
<td>Subject: NnAg/Fro</td>
<td>0.99</td>
</tr>
<tr>
<td>Items: (intercept)</td>
<td>0.07</td>
</tr>
<tr>
<td>Items: Ag/Fro</td>
<td>0.15</td>
</tr>
<tr>
<td>Items: Ag/Beh</td>
<td>0.70</td>
</tr>
<tr>
<td>Items: NnAg/Fro</td>
<td>0.36</td>
</tr>
<tr>
<td>Items: NnAg/Beh</td>
<td>0.29</td>
</tr>
</tbody>
</table>

\(^1\) Highest correlation between fixed effects, \(r = 0.38\).

To investigate whether the exposure to other experimental conditions may have affected these results, despite the blocked design, we conducted an additional exploratory analysis on block 1 only in order to isolate the effect of first exposure to each Entity condition. Figure 33 shows the proportion of non-self descriptions in block 1 only. In a mixed logit model, we included the deviation-coding numeric predictor comparing each individual Entity - Location condition against the No Entity baseline as fixed effects, and the full random structure justified by the design (by-subjects, this only included random intercept). Only the Agent-Behind condition triggered significantly more non-self perspective descriptions compared to
the No Entity baseline ($\beta = 1.54$, S.E. = .55, odds-ratio= 4.65, 95% CI [1.60; 13.57], $p = 0.005$), all other comparisons were not significant ($ps > .1$).

![Figure 33](image-url)

*Figure 33* Observed proportions of non-self descriptions in Block 1 of Experiment 6. Error bars represent non-parametrically bootstrapped (BCa) 95% confidence intervals on the participant-wise conditional means.

### 7.3.4 Discussion

When participants described Agent, Non-agent, or No Entity scenes in a blocked presentation, they were consistently more likely to adopt the non-self perspective in the Agent-Behind condition compared to the Agent-Front condition (as in Experiment 5). However, they showed no such tendency in the Non-Agent conditions. These findings therefore provide further support for the perceived intentional account, and suggest that in Experiment 5, participants transferred intentionality and orientation from agents (people) to non-agents (plants). It must be noted, however, that in Experiment 6 we did not find that participants were more likely to describe the scene from a non-self perspective in the Agent-Behind condition compared to the No Entity baseline. This difference was significant once we limited the analysis to Block 1 alone, when participants first experienced the task and were exposed to only one Entity condition. The fact that the effect of Agent-Behind vs. No
Entity faded away when the data from the three blocks were collapsed into a single analysis may be due to the variability in the effect induced by the exposure to other experimental conditions, or because of loss of motivation or decreased engagement with the task in subsequent blocks.

7.4 EXPERIMENT 7

Taken together, Experiments 5 and 6 suggest that speakers’ tendency to adopt a non-self perspective to describe the location of objects in a scene can be affected by the presence of an agentive entity that has a different perspective and that evinces no current action or intention to act. In both experiments, we found an increased likelihood of adopting a non-self perspective in the Agent-Behind conditions (where the agent could potentially see and act on the objects whose location was being described) than in the Agent-Front conditions (where the agent could not), when this agent’s perspective represented a contrasting perspective to the speaker. Our results are consistent with the perceived intentional account, in which the speakers’ perception of this entity’s ability to see or act on objects is relevant during simulation and is functional to non-self perspective taking. They are not consistent with the non-intentional account (which assigns no role to agency), or the actional intentional account (which requires the agent to evince action or an intention to act).

In Experiment 7, we tested the perceived intentional account further by concentrating on agentive entities, and manipulating both their location with respect to the objects (Front vs. Behind; Figure 34a,b vs. 4c,d) and their orientation with respect to the participant’s orientation (Aligned vs. Misaligned; Figure 34b,d vs. 4a,c). We tested whether the effect found in the previous two experiments, with speakers being affected by another agent's perspective if this latter could see the objects, is limited to situations where the other's perspective is opposite to one's own (and so speakers produce more non-self perspectives when the “misaligned other” can see the objects, as in Experiment 5 and 6) or it extends to situations where the other's perspective is the same as one's own (if so, speakers should produce fewer non-self perspectives when the “aligned other” can see the objects).

If a speaker’s choice of perspective is affected by the perception of another agent only when this agent has both the ability to see and act on the relevant objects and a different
perspective to the speaker's own, then we would predict more non-self descriptions when the agent had a different orientation to the participant and was facing the objects (and so could see and potentially act on them; Figure 34c) than when the agent had a different orientation to the participant but was not facing the objects (and so could not see and potentially act on them; Figure 34a). However, there should be no difference in non-self descriptions when the agent had the same orientation (and so the same perspective) as the participant, irrespective of whether they were facing or not facing the objects (Figure 34b, d). We should thus expect an interaction between the agent's Location and Orientation.

Instead, under a more fully intentional explanation, the perspective of the other agent should influence a speaker's choice of perspective even when their perspective matches the speaker's if the agent can see the objects. Given that their perspectives match, however, this influence would amount to fewer non-self descriptions when the (aligned) agent is facing the objects (and so could see and potentially act on them; Figure 34d) than when the (aligned) agent is not facing the objects (and so could not see and potentially act on them; Figure 34b). We should thus expect a main effect of Location and Orientation but no interaction.

7.4.1 Methods

7.4.1.1 Participants

288 participants took part in the online experiment.

7.4.1.2 Materials

We created 12 experimental items. As in Experiment 5, an item corresponded to two experimental objects with no clear front/behind or left/right. For each item, we created two sets of five photograph scene: one set for each possible position of the two objects on the table (see Experiment 5); five photographs corresponding to No entity, and the four unique combinations of Location and Orientation (see Figure 34).

We then created 12 lists, each comprising 24 experimental and 16 filler trials. Overall, each list contained 8 No Entity conditions and 16 Person conditions, 4 for each combination of Location and Orientation. In each list, each item appeared twice but never in the same condition, and participants were asked about the position of each individual object only once.
Figure 34 Example of a set of photographed scenes (item: candle/pineapple) used in Experiment 7: (a) Front Misaligned; (b) Front Aligned; (c) Behind Misaligned; (d) Behind Aligned.

7.4.1.3 Design

As in the previous experiments, participants saw photos of scenes in which two objects were located on the opposite sides of a table. They described the spatial relation between the two objects by answering the question: *On which side of X is Y?*. We manipulated: (i) whether an agentive entity (i.e., a person) was present in the scene (Entity: No Entity vs. Agent), (ii) whether the agent was behind or in front of the table (Location), (iii) whether the agent’s orientation towards the scene was aligned with the speaker’s or not (Orientation).

The study used an incomplete factorial design with No Entity as control group and a 2×2 within-subject within-item design for Entity Person with Location (Behind vs. Front) and Orientation (Aligned vs. Misaligned) as factors. The experiment was identical to Experiment 5 except for the following details.

7.4.1.4 Procedure

The procedure was as in Experiment 6.
7.4.2 Data analysis

7.4.2.1 Coding and exclusion criteria

35 participants (12%) had taken part in one of our previous experiments and were excluded from the analysis. 26 participants (9%) did not have English as first language and other 4 (1.4%) provided no response; they were also excluded. 17 descriptions contained explicit reference to both perspectives; as they only amounted to 0.25% of all data points, they were removed. In 46 trials (0.67%), no perspective was taken (e.g., “the pear is at the opposite end of the table”), and in other 36 (0.52%) no response was provided either due to participants or problems with non-uploading pictures. They were all excluded. We further excluded from the analysis 105 participants (36.5%) who never produced a non-self perspective and 2 others (0.7%) who never produced a self-perspective. Data from 114 Subjects, those who produced descriptions from both perspectives at least once, were included in the analysis for a total of 2713 data points (897 in No Entity, 452 in Behind/Aligned, 454 in Front/Aligned, 455 in Front/Misaligned and 455 in Behind/Misaligned). The analysis predicts the likelihood to produce a non-self spatial descriptions (\(non-self = 1; self = 0\)).

7.4.3 Results

Participants used a non-self perspective to describe the spatial location of the objects on 18% of trials (see Table 22 and Figure 35).

<table>
<thead>
<tr>
<th>Orientation</th>
<th>Location</th>
<th>Perspective Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>self</td>
</tr>
<tr>
<td>Misaligned</td>
<td>Front</td>
<td>362</td>
</tr>
<tr>
<td></td>
<td>Behind</td>
<td>338</td>
</tr>
<tr>
<td>Aligned</td>
<td>Front</td>
<td>394</td>
</tr>
<tr>
<td></td>
<td>Behind</td>
<td>376</td>
</tr>
<tr>
<td>No Entity</td>
<td></td>
<td>758</td>
</tr>
</tbody>
</table>
A logit mixed model with Location, Orientation and their interaction as fixed effects (see Table 23) revealed a main effect of Orientation: Participants were more likely to produce non-self descriptions when the agent held an opposite orientation to their own compared to when it had the same orientation ($p < 0.05$). More interestingly, the significant main effect of Location ($p < 0.05$) indicates that, regardless of whether the agent’s orientation matched or not the participant’s, participants produced more non-self descriptions in the two behind (vs. front) conditions. This means that participants’ choice of perspective was sensitive to whether the agent could or could not see and act on the objects in the misaligned and the aligned Orientation conditions. There was no interaction between Location and Orientation ($p > .5$), which was confirmed by a BIC-based Bayes Factor analysis leading to a factor of $e^{(1701.5 - 1694.3)/2} = 36.6$ in favour of the no-interaction model (corresponding to a posterior probability of 0.97).
### Table 23
Summary of the mixed logit model with Location (Behind = .5/ Front = -.5) and Orientation (Misaligned = .5/ Aligned = -.5) as predictors for Experiment 7 (N=1816)^1

<table>
<thead>
<tr>
<th>Predictors (fixed effects)</th>
<th>Parameter estimates</th>
<th>Wald's test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log-odds ($\beta$)</td>
<td>S.E.</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.89</td>
<td>0.14</td>
</tr>
<tr>
<td>Orientation</td>
<td>0.49</td>
<td>0.23</td>
</tr>
<tr>
<td>Location</td>
<td>0.46</td>
<td>0.19</td>
</tr>
<tr>
<td>Orientation:Location</td>
<td>-0.22</td>
<td>0.38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Explained standard deviation estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects: (intercept)</td>
<td>1.15</td>
</tr>
<tr>
<td>Subjects: Orientation</td>
<td>1.25  0.34</td>
</tr>
<tr>
<td>Subject: Location</td>
<td>0.17  -0.56  0.59</td>
</tr>
<tr>
<td>Subject: Orient:Loc</td>
<td>0.44  0.94  0.02  -0.80</td>
</tr>
<tr>
<td>Items: (intercept)</td>
<td>0.00</td>
</tr>
<tr>
<td>Items: Orientation</td>
<td>0.29</td>
</tr>
<tr>
<td>Items: Location</td>
<td>0.00</td>
</tr>
<tr>
<td>Items: Orient:Loc</td>
<td>0.00</td>
</tr>
</tbody>
</table>

^1 We forced independence between the by-item random effects following perfect correlations between all the by-item random effects. This did not affect the significance of the fixed effects. All correlations between fixed effects r < |0.25|.

To verify whether both Conditions in which the agent could see the objects triggered a different extent of non-self perspectives than chance, we performed an additional mixed logit analysis comparing the No Entity baseline with the individual four levels of the 2×2 design. We tested for this by creating one factor representing each combination of Location and Orientation, and then converting this factor to a deviation-coding numeric representation. The model summary is presented in Table 24. Only the Behind-Misaligned condition triggered significantly more non-self perspective descriptions compared to the No Entity baseline (p < 0.01). The front-aligned condition did not significantly reduced the likelihood of non-self perspective compared to the baseline, even if data go in that direction ($\beta$ = -0.49, p = .12).
### Table 24

Summary of the mixed logit model comparing the No Entity baseline and the four unique combinations of Location and Orientation via deviation coding (N=2713)

<table>
<thead>
<tr>
<th>Predictors (fixed effects)</th>
<th>Parameter estimates</th>
<th>Wald's test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log-odds ($\beta$)</td>
<td>S.E.</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.92</td>
<td>0.13</td>
</tr>
<tr>
<td>Misaligned/Front</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>Misaligned/Behind</td>
<td>0.56</td>
<td>0.22</td>
</tr>
<tr>
<td>Aligned/Front</td>
<td>-0.49</td>
<td>0.32</td>
</tr>
<tr>
<td>Aligned/Behind</td>
<td>0.21</td>
<td>0.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Explained standard deviation estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects: (intercept)</td>
<td>1.12</td>
</tr>
<tr>
<td>Subjects: Misal/Fro</td>
<td>0.76 0.32</td>
</tr>
<tr>
<td>Subject: Misal/Beh</td>
<td>0.89 0.44 0.99</td>
</tr>
<tr>
<td>Subject: Alig/Fro</td>
<td>0.69 0.18 -0.84 -0.77</td>
</tr>
<tr>
<td>Subject: Alig/Beh</td>
<td>0.46 -0.60 -0.82 -0.87 0.64</td>
</tr>
<tr>
<td>Items: (intercept)</td>
<td>0.00</td>
</tr>
<tr>
<td>Items: Misal/Fro</td>
<td>0.00</td>
</tr>
<tr>
<td>Items: Misal/Beh</td>
<td>0.00</td>
</tr>
<tr>
<td>Items: Alig/Fro</td>
<td>0.46</td>
</tr>
<tr>
<td>Items: Alig/Beh</td>
<td>0.00</td>
</tr>
</tbody>
</table>

1 We forced independence between the by-item random effects following perfect correlations between all the by-item random effects; significance results were unaffected. Highest correlation between fixed effects, $r = 0.44$.

#### 7.4.4 Discussion

When participants described Agent or No Entity scenes, they were more likely to adopt the non-self perspective when the scene contained an Agent that had a different perspective from their own, and that could see and act upon the objects (but evinced no current action or intention to act), than when it contained no entity. They did not show a greater tendency to adopt a non-self perspective when the scene contained an Agent that had the same...
perspective as their own, or had a different perspective but could not see and act upon the objects, than when it contained no entity.

However, participants’ choice of perspective was sensitive to whether the agent could or could not see and act on the objects, both when this agent shared and when it did not share participants’ perspective. When participants perceived an agent that had a different perspective to their own, they produced more non-self descriptions when this agent showed the ability to see and act on objects (Figure 34c) than when it did not, as the agent’s perspective corresponded to the non-self perspective corresponded. When participants perceived an agent that shared their perspective, they produced fewer non-self descriptions when the agent could see the objects (Figure 34b) than when it did not, presumably because the agent’s perspective corresponded to (and hence reinforced) the participant’s self perspective.

7.5 General Discussion

When people describe the location of objects, they can do so from their own perspective or another perspective. In our experiments, we have shown that their choice of perspective is affected by the presence of another person in the scene that they are describing, even though they are not communicating with that person. When speakers described the spatial relationship between two objects in photographed scenes, they were more likely to abandon their self perspective when the scene included a person (i.e., an agent) with a different perspective to their own that could see the objects, compared to a scene that contained no other entity. This tendency occurred even though the agent was not acting on (or showed an intention to act on) the objects. However, there was no such tendency when the agent could not see the objects (Experiments 5, 6 and 7). Nor was there a tendency to abandon a self perspective when the scene included a non-agent (i.e., a plant, of a similar size/shape to the agent) and speakers had not previously described scenes including an agent (Experiment 6). Additionally, speakers were more likely to abandon their self perspective when the scene included an agent with a different perspective who could see the objects, compared to a scene that contained no other entity, but not when the scene included an agent with the same perspective (Experiment 7).
Our results suggest that during spatial language use, speakers’ choices can be affected simply by the presence in the environment of an agentive entity with the potential to act, even if that agent is not currently carrying out an action or evincing an intention to act. These results are inconsistent with the actional intentional account, whereby by the presence of an agent who is currently acting or clearly intending to act is a required trigger to non-self perspective. They are also not consistent with the non-intentional account, whereby speakers’ choice of spatial perspective is affected by featural or functional aspects of the environment that are independent of the availability of another agent’s self perspective. Under that account, speakers’ tendency to adopt a non-self perspective when describing the location of objects should have been uniformly affected by the presence of an entity with discernible axes that could provide a spatial reference point (e.g., front/behind; left/right), irrespective of whether it could act on the objects. In contrast, our results showed that non-self perspectives were triggered only when an agent (which has discernible axes regardless of its orientation) was able to see (hence potentially act on) the objects. Instead our results are consistent with the perceived intentional account, whereby the availability of an agent with a different perspective and the potential to act can trigger a non-self perspective.

To explain our results, we propose that speakers incorporated the agent’s location and orientation, and more specifically, the agent’s visual perspective (i.e., what they could and could not see), into their representation of the objects’ spatial arrangement. That is, when the agent was oriented in such a way that the objects were in their visual field, the speaker was induced to mentally simulate the scene from the agent’s perspective. As a result, when participants’ (self) perspective did not match the agent’s perspective, they sometimes shifted to a non-self perspective when describing the location of the objects. Our proposal is compatible with research on non-linguistic spatial cognition showing that another person’s visual perspective can interfere with a participant’s performance on spatial (Zwickel & Müller, 2010) and non-spatial (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010) target detection tasks, so long as that perspective appears salient.

We propose a simulation account in which apprehending the visual perspective of another person makes speakers aware of a spatial perspective other than their own. This awareness can induce them to sometimes adopt that perspective to describe spatial relationships. Such
mechanisms seem to be linked to agency but are independent of more explicit action-related considerations (e.g., action understanding or discernible intentionality).

Our proposal is congruent with previous findings that people can easily and quickly process what another person can or cannot see, without resorting to effortful mental representations (Kessler & Rutherford, 2010; Michelon & Zacks, 2006; Surtees et al., 2013b). The participants in our experiments seemed to have relied on this ability and quickly processed whether the agent could or could not see the objects; this information then affected whether they simulated the agent's spatial perspective in the scene. This is important, as processing what an agent can or cannot see before simulating their position in space can prevent us from unnecessary "costly" simulation (cf. Johnson & Demiris, 2005).

We cannot be sure how consistently the observed agent’s visual and action affordance triggered participants to simulate the agent’s perspective. Overall, people produced non-self descriptions infrequently, showing a strong tendency across the three experiments to produce descriptions that adhered to their self perspective even when the person in the scene had an opposite perspective and could see the objects (self descriptions: 77% in the agent-behind condition across the three studies). However, the perspective that people adopt when describing space is not necessarily the perspective through which they first conceive that space, so analysis of their descriptions is not directly informative about the spatial perspective that they initially take.

One possibility is that participants usually simulated the non-self perspective in the presence of a relevant environmental trigger (i.e., the presence of an agent) but only seldom translated this perspective into the corresponding linguistic expression (i.e., “left”/“right”). A related possibility is that they usually simulated both perspectives in parallel, but that the self perspective most frequently won the competition to be linguistically realized. Verbal left/right judgments from positions other than the self have been shown to require more effort than equivalent motor responses (Kessler & Thomson, 2010), probably due to the underlying ambiguity of “left” and “right” terms (Coventry & Garrod, 2004). Hence, participants may have avoided that extra effortful step. The fact that there was no incentive to do so in the task may have exacerbated this tendency. It is also possible that what an
agent can see may not always be a sufficiently strong enough cue to the agent’s relevance and perspective to trigger a non-self perspective in the speaker.

Our results confirm previous findings from non-linguistic spatial cognition that the experiential context can strongly affect the spatial perspective that people adopt (cf. Surtees et al., 2012; Clements-Stephens et al., 2013). In our study, inanimate entities without clear left and right axes were able to trigger non-self descriptions when presented intermixed with agentive entities (Experiment 5). They were not able to do so when they were presented alone (Experiment 6). Future studies should carefully account for this experiential factor when testing for agency- or human-related effects, and need to incorporate non-agentive control conditions.

Our results are not consistent with an account based on agency-independent object-centred spatial coding, in which participants used the additional entity as a source of spatial coordinates, that is used the entity's left/right axes to parse space without simulating the alternative perspective (Heyes, 2014). First, objects without clear intrinsic left/right axes (plants) triggered non-self descriptions to the same extent as agents in Experiment 5. Second, an intrinsic-axes mapping explanation would wrongly predict an effect of agency irrespective of the location of the agent. It would not predict that participants would produce more non-self descriptions when the agent appeared behind the table than in front of the table, as we found in Experiments 5 and 6, because the directional features of the human figures do not change between these two conditions. The pattern is even more striking for Experiment 7, where the production of non-self descriptions was modulated by the agent’s orientation with respect to the objects (i.e., whether the agent could or could not see the objects), even when their left/right axes matched those of the participant.

However, these considerations do not rule out the possibility that intrinsic-axes mapping mechanisms may have been implicated in previous studies in which the scenes to be described included only the located object and the person. This type of setting may in fact have induced participants to use the only other available entity (i.e., the person) as a source of spatial encoding for the located object. Participants may have felt encouraged to use the person’s body axis as a basis for their spatial descriptions, especially when the person’s pose (e.g., head turned, arm stretched) created asymmetries that made the person’s left/right axes
more salient. Such an explanation could account for why previous studies using this paradigm found that pictures including a person engaged in action but not in a neutral pose triggered more non-self descriptions than a baseline picture that did not include a person (e.g., Mazzarella et al., 2012). We therefore suggest that simulation and intrinsic-axes mapping mechanisms may not be mutually exclusive, but may instead exert additive effects.

Overall, our findings are consistent with the view that spatial perspective taking in language is linked to speakers processing others’ representation of a shared environment, and in particular, what information others may be ignorant or aware of. This conclusion is strengthened by previous findings that individuals tend to ignore stimuli that another person ignores (e.g., Frischen, Loach, & Tipper, 2009). It is also in line with existing evidence that knowing that another person is looking at the same stimuli affects how people encode spatial information to solve a target detection/inhibition task (Böckler et al., 2011; Böckler & Zwickel, 2013). In our experiments, a more basic form of “shared attention” may be implicated. The person in our photographed scenes never directed their gaze at the objects, but in some sense experienced the same scene as the participant (even if from a different perspective, and presumably at a different point in time). Future studies should try to pin down whether, if shared attention is indeed modulating the effect, this can only be triggered if the person is represented as physically present in the observed environment or, instead, can also be induced by the mere belief that someone else is attending the same environment at the same time.

7.6 CONCLUSIONS

In conclusion, our results suggest that even when people are not motivated by communicative goals, they sometimes spontaneously adopt a non-self perspective to describe space. Crucially, and in line with the perceived intentional account of spatial perspective in non-interactive language, this spontaneous shift is linked to simulation processes induced by the presence of another agent who is perceived as able to see and act on the jointly-attended scene. Our results also show that this agent does not have to be currently engaged in an action or show any clear intention to act in order to trigger a non-self perspective in the speaker. Finally, we showed some evidence that when perceived as relevant, where relevance is determined by being able to see and act on the objects, the perspective of an
observed agent may also reinforce self perspective in situations where it matches the speaker's.
Chapter 8

THE EFFECT OF SHARED EXPERIENCE ON PERSPECTIVE TAKING IN SPATIAL LANGUAGE

8.0 CHAPTER OVERVIEW

In Chapter 7, we saw that speakers' choice of perspective to locate objects in space can be sensitive to the vantage point of another person in the environment, under the condition that the person is also potentially visually aware of the relevant objects. In this chapter, we look more closely at the effect that the awareness of shared perceptual experience may have on spatial perspective in language production. We explore the possibility that the influence of another person's vantage point may not be limited to cases in which the spatial relation between the other person and the environment can be directly perceived (social stimuli) but might extend to situations in which the spatial relation can only be assumed by the context. In two online experiments, we investigated whether the mere belief that another person was simultaneously engaged in the same task with the same stimuli, yet from a different vantage point, modulated the perspective that participants would adopt to verbally describe the stimuli. We further explored whether the effect could extend to the belief about a crowd of co-experiencers. We found that the belief of shared experience reinforced self-perspective taking when the other person’s and participant’s perspective were the same; the effect did not extend to crowds.

8.1 INTRODUCTION

You are attending a tennis match and are not very pleased with the seat you managed to get, right on the baseline behind the player you support. You receive a text from a friend that she is also at the match. Funnily enough, it turns out she is seated right opposite you on the other side of the court. Would this belief affect your experience of the match and, in particular,
how you would indicate to a third person the side of the court on which your player is now standing and calling a ball out?

Shared experience is the awareness that someone else is attending to the same things or events and is engaged in the same task at the same time as you are (Garriy Shteynberg, 2015). As we saw in Chapter 7, speakers can be influenced by another person's vantage point when producing spatial descriptions of visual scenes if the relevant objects were located within the person's visual field. We argued that the effect could be due to the speaker simulating the other person's spatial perspective when that person’s role in the scene was perceived as relevant, where 'relevance' was determined by visual affordance (i.e., the person could see the objects). An alternative yet compatible explanation is that people were induced to process, and sometimes adopt, the other's perspective not so much by the fact that the other person could see the objects, but rather by the fact that the objects on which the participants were focusing their attention could be seen by the other. In other words, the triggering factor was the awareness that they both could simultaneously see the objects: that their attention was shared.

Evidence that people spontaneously consider what another person is attending to in a way that affects how they process and talk about space, comes from individual-task paradigms in which participants described the spatial layout of scenes that also included a person. In Tversky & Hard (2009), speakers were likely to describe an object's location from another person's opposite perspective if the person was perceived as paying attention to the object (by looking at or reaching for it), even if taking the person's perspective was not required by the task. In addition, what another person can see has been found to interfere with the judgment of what one oneself can see (number of dots detection task, Samson et al. 2010; but see Santiesteban et al., 2014, for an alternative explanation).

Note that in all these studies, the other person was a physical element of the environment to be described and participants had direct perceptual access to the person's spatial relation to the other objects. What these findings may thus show is that speakers' choice of spatial perspective can change depending on whether they are processing social versus non-social stimuli. In this chapter, we asked whether a similar effect can be found if this social element is moved from the stimuli (i.e., the observable presence of a person) to the experimental
context (i.e., the assumption of co-attention). Would the mere belief that another person is jointly attending to the same environment, yet from a different perspective, modulates the perspective that a speaker adopts to locate objects in that environment?

There is, to our knowledge, no direct evidence that the awareness of shared experience can affect a person's spatial language perspective in situations that do not require any communicative requirements, whether these requirements are real (e.g., Schober, 1995) or only imagined (e.g., Schober, 1993). The explicit communicative nature of previous psycholinguistic studies, in fact, might have required speakers to adjust to their addressee's perceptual and spatial relation to the world in order to fulfil this communicative demand (i.e., make sure that they understood each other)\textsuperscript{44}. Instead, we are interested in isolating the effects from the demand of social interaction, and see whether similar adaptations occur if people merely believe that they are actively experiencing something at the same time as someone else.

8.1.1 The belief of shared experience and its possible relevance for spatial language perspective taking

Recent studies have started investigating how another person's spatial relationships within a context of shared experience (but lacking interaction or communication) may affect how people attend to or cognitively process the jointly-attended stimuli (e.g., Böckler & Zwickel, 2013). Böckler, et al. (2011) had participants perform a handedness recognition task at the same time as another person who was sitting opposite them, thus attending to the same stimuli from an opposite vantage point. Participants saw images of two hands in succession and judged whether the second hand, which was rotated to different degrees with respect to the first one, was the same hand (left vs. right) as the first hand. On relevant trials, the first hand was always presented at 0° degree of orientation, matching participants' own hand. Assuming that handedness is judged by performing a mental rotation of one's own hand, participants were predicted to be slower at responding to hands of higher degrees of rotation. Indeed, when the other person kept their eyes closed (no shared experience), participants

\textsuperscript{44} Note that we are not making any claims as to whether these partner-oriented adjustments are due to speakers genuinely modelling their partner's needs (e.g., Clark, 1996) or instead they emerge through adaptation driven by the interaction (see Barr, 2014).
showed the expected processing difficulty effects at higher degrees of rotation. However, there was no sign of processing difficulties in the shared-experience condition, when the two participants simultaneously performed the task. This suggests that participants might have benefitted from processing the stimuli from the other person's perspective (or, at least, used the other person's body axes to map and process stimuli) on condition that they thought that the other person was also actively engaged. As in the individual-task settings discussed above, also in this study the other person was physically present and so provided a strong cue to their own spatial reference. However, the fact that the effect only occurred when both participants actively attended to the stimuli together hints toward an important modulatory role for shared experience. It remains nonetheless unclear whether similar interpersonal effects may extend to non-manual responses, like verbal descriptions of space.

More generally, there is good evidence that simultaneously attending to the world together with another person affects how we perceive the world and makes us spontaneously consider the perceptual relations between that person and the world. Richardson et al. (2012) tracked the eye gaze of participants presented with neutral, positive and negative images, while they were told that another person was simultaneously looking at the same (vs. a different) set of stimuli. They found that this belief shaped what people focused their attention on: they looked longer at the images that they presumably “thought” were salient for the other person (i.e., the negative images). Other experimental findings have been used to suggest that people cannot help but shift cognitive resources and engage in greater elaborative processing towards objects of shared attention (Garriy Shteynberg, 2010). For instance, people were better at remembering stimuli that either they or their co-participant had to respond to (individually) in a previous task compared to the stimuli that they both had to ignore (Eskenazi, Doerrfeld, Logan, Knoblich, & Sebanz, 2013). The awareness of being embarked on the same endeavour has also been found to shape object evaluation, with people amplifying their appreciation/depreciation of the shared experience (e.g., the taste of a piece of chocolate; Boothby, Clark, & Bargh, 2014), and to increase their affective responses to the shared situation, like smiling more (Fridlund, 1991; see also Shteynberg & Galinsky, 2011), compared to individual experiences. This is not only true for adults. Children tend to more easily recognize objects that they have co-attended to with a caregiver than objects they have attended to alone (e.g., Tomasello, Carpenter, Call, Behne, & Moll, 2005).
Interestingly for our study, these effects seem to depend not only on people attending to the same stimuli together but also on them assuming that they are simultaneously engaged in the same task. Going back to Richardson et al.’s (2012) eye-tracking study, people were no longer sensitive to what image might have attracted the other person’s attention if they knew that, although looking at the same set of images, the other was engaged in a different task (memory vs. search task). Similarly, Gambi, Van de Cavey and Pickering (2015) found that participants took longer to name pictures if they believed that another person was simultaneously naming pictures too, even if they could not hear or see each other. Interestingly, the effect was independent of the actual content of each other’s utterance (e.g., it occurred regardless of whether the other named the same or a different picture) but disappeared if the other person was engaged in a categorisation task or remained silent.

Equally relevant for our investigation is the fact that people have been shown to be influenced by others’ focus of attention even in the absence of physical co-presence, when they did not have direct perceptual access to each other’s actions but merely believed that another person was carrying out the task with them (e.g., Atmaca et al., 2011; Gambi et al., 2015, see also Ruys & Aarts, 2010 for similar results) and when they could see the other but did not receive feedback about their responses (Vlainic, Liepelt, Colzato, Prinz, & Hommel, 2010). This is also true for online settings where people could only assume the co-attention of the other person (e.g., Shteynberg, Hirsh, Galinsky, & Knight, 2014). This is important in that it shows that the simple knowledge about what another person is about to do or is doing can get integrated into how we perceive and process the jointly-attended world (yet see Sellaro, Treccani, Rubichi, & Cubelli, 2013 for conflicting results). As in real life, in fact, shared experience does not require the simultaneous attention to be real. You can believe you are co-experiencing a tennis match with your friend even if your friend was actually delayed in the traffic and is not there yet, or she is there but is busy taking a selfie of herself with the tournament mascot.

The findings reviewed so far are compatible with a broader body of evidence from “joint action” research that when performing a task with someone else, people cannot help but cognitively represent information about the other person’s role in the joint activity (e.g.,
Sebanz, Knoblich, & Prinz, 2003; Sebanz et al., 2005), even if this information is not relevant, or is even counter-productive, to one’s own role in the task.

Most evidence in this regard comes from so called joint stimulus-response reaction-time paradigms in which a two-choice task is distributed between two people, with each person responsible for one action alternative (e.g., participant A responds to green stimuli with right key, participant B responds to red stimuli with left key, Sebanz et al., 2003). Several studies (e.g., Hommel, Colzato, & van den Wildenberg, 2009; Tsai, Kuo, Hung, & Tzeng, 2008; Welsh et al., 2013) have demonstrated how the presence of the other active individual interferes with one’s own stimuli processing and task performance; people show in fact the same compatibility and interfering effects that occur when one single individual is in charge of both action alternatives by themselves. People respond faster when the stimulus’ and the response’ position are the same (e.g., right button in response to a green stimulus displayed on the right of the screen) compared to when they are opposite (e.g., right button / green left-displayed stimulus); this occurs when people act alone and are responsible for both responses (i.e., left/right buttons) and both stimuli (e.g., red/green) as well as when two people act together and each is only responsible for one response (e.g., right button) to one assigned stimulus (e.g., green).

In line with the empirical evidence reviewed so far, findings from joint-action research have assigned a central role to active shared attention (Sebanz & Knoblich, 2009); that is, people must have attentional access to each other’s stimuli and be actively engaged in the jointly-attended environment for such effects to occur. For example, using a joint Navon task in which two participants attended to the same stimuli together (big letters constituted by smaller letters), Böckler, Knoblich and Sebanz (2012) found that people responded slower and were less accurate when the other had to focus on and respond to a different dimension of the stimulus (small letters vs. big letters) than when they focused on and responded to the same dimension. Crucially, participants were attending jointly to the same stimuli but, and this is important for our study, the effect held even if participants looked (or thought they were looking) at two different instances of the same stimulus; a bit like when people cheer for the same tennis player in a match but are watching the match on different TV screens. Thus, mutual visual access (or belief thereof) is not contingent upon the same stimulus token.
8.1.2 Under which mechanisms could shared experience affect perspective in spatial language?

The evidence reviewed above suggests that how people perceive and act upon the world is affected by their knowledge that someone else is simultaneously paying attention to the same things and is concurrently engaged in the same task. On the one hand, people seem to process the other person's perceptual perspectives on the jointly-attended environment and integrate these into their own processing of the world (Richardson et al., 2012). As a result, people's affective and cognitive responses are amplified (Shteynberg & Galinsky, 2011), and even lower-level cognitive mechanisms like memory and spatial encoding or attention are shaped by the assumed focus of attention of the other (Böckler et al., 2012). On the other hand, the other person's actions and intentional stance towards the jointly-attended objects generate response conflicts if different from one own, thus suggesting that they get integrated into one's own action planning and execution (Sebanz et al., 2003). Importantly, these social-influence effects happened independently of whether the other person was actually there, or their presence was only imagined or inferred. The belief of synchronous co-attention and mutual perceptual access to the relevant objects seem instead to be essential prerequisites.

How can this be informative as to whether and how the perspective a person adopts to locate objects in space is shaped by the belief that another person is concurrently attending to the same space and doing the same task yet from a different vantage point? One possibility is that sharing an experience in real time with another person (e.g., passively observing images or being engaged in a complementary task) may spontaneously induce people to consider each other's feelings, thoughts, perspective and goals towards what is shared – a bit like watching a tennis match with a friend may make her more prominent to you than when you are watching the match and she is involved in a different activity (e.g., you may notice Federer's two pairs of socks because you know she has a big crush on him or smile when Sharapova scores because your friend is a big fan). For instance, it has been suggested that once we are aware of co-attending with another person, we naturally integrate our co-experiencer and their (perceptual, spatial or intentional) relationship to the world into our own experience (Boothby et al., 2014).
This is in line with the action co-representation hypothesis to joint action which predicts that, when made aware about the role of another person in a joint task, people cannot help but represent the other's co-task (i.e., what the other person has to respond to and how), and integrate it into their own action plans (Milanese, Iani, & Rubichi, 2010; Sebanz, et al., 2005). As a consequence, interference effects at the level of action-plan selection (slower responses) occur when stimuli contain features that are relevant for both one's own and the other's response (e.g., Gambi et al., 2015; Atmaca et al., 2011; Böckler et al., 2012) or that cue to the other's response (e.g., Milanese). As we saw, however, action co-representation is not always detrimental (e.g., Sellaro et al., 2013) but may lead to facilitation effects depending on the task demand (Böckler et al., 2011), previous experience and experimental context (Milward, Kita, & Apperly, 2014).

Less is known about the extent to which these effects may be due to higher-order mentalising mechanisms; that is, are people representing the intention/goal of the other person and integrating them into their cognitive representation of the joint task? Some support for a role of mentalising comes from studies which found joint-action effects only when the co-participant acted intentionally but not when a machine produced the action (Atmaca et al. 2011, but see Dolk et al., 2011; Stenzel et al., 2014 for different results). Also, findings showed joint-action effects in older (above 4) but not younger children (Milward et al., 2014), nor in patients who failed a ToM task (Humphreys & Bedford, 2011).

An alternative yet related explanation from shared-attention research suggests that people might be automatically be driven to prioritize and orient their attention towards objects and events to which other people, and especially people to whom they feel similar, are attending; a mechanism called “social tuning” (Shteynberg & Galinsky, 2011; Shteynberg, 2015). According to Shteynberg (2010), for instance, shared attention could spontaneously induce a first-person-plural perspective; a we mode that, without mentalising, would allow us to “perceive the self and the other as a unified agent with a singular attentional focus” (p. 581). This would explain, for instance, why shared-attention effects may happen among an indistinct crowd (like in a stadium; see also Boyd & Richardson, 2009).

In sum, according to all these hypotheses, for another person's concurrent action and attention to impact another individual's cognitive processes, the person involved must be
aware of some perceptual experiences being shared and appreciate that some other agent is intentionally co-acting (Sebanz & Knoblich, 2009; see also Wenke et al., 2011 for a slightly different hypothesis).45

8.1.3 The current study

The studies reviewed above demonstrate that people are highly sensitive to what others are focusing their attention on and doing, if they have this information. In our study, we address the question as to whether this information also affects the perspective they adopt to describe space.

We attempted to test whether knowing the vantage point of a person who is concurrently attending to the same stimuli and carrying out the same task shapes the perspective that people adopt to describe the spatial relationships between the jointly-attended stimuli (similarly to what has been found when people could directly perceive the other person's vantage point). In other words, does the effect depend on people observing the spatial relations between the other person and the objects or could it be induced by the mere awareness of these relations?

If, as shown above, people acting at the same time represent the other's task and focus of attention even when the other is not physically there, then it may be that people in our scenario will be similarly affected by the other person's vantage point and adopt their perspective when carrying out the task.

In two online experiments, we presented participants with photographed scenes of two objects on a table and asked them to locate one object with respect to the other. Participants were made to believe that another person would be doing the experiment at the same time as them. Participants were also cued about whether the other person would have the same vantage point on the scenes as them (i.e., they were looking at the same scene taken from the same side of the table) or, instead, had an opposite vantage point (i.e., the photographs that

45 Note, however, that this “socially rich” approach has been challenged by findings that joint stimulus-response compatibility effects can be triggered by object-generated events (e.g., a ticking metronome, Dolk, Hommel, Prinz, & Liepelt, 2013), thus suggesting that the effects may be due to attention-orientating (Dolk et al., 2014) and/or spatial-coding mechanisms (e.g., Guagnano, Rusconi, & Umiltà, 2010).
the co-participant was seeing were taken from the opposite side of the table). We investigated whether the belief about the vantage point of the co-experiencer would affect the spatial perspective that people adopted in their task responses. In Experiment 9, we also varied the number of co-experiencers (one versus many) to investigate whether the effect could be enhanced by a crowd of five other people.

People briefly interacted with the co-participant (in reality a pre-set computer program) at the beginning, between the two experimental blocks, and at the end (in the crowd condition of Experiment 9, the interacting co-participant was “randomly chosen” among the five). However, this (fake) exchange was restricted to a side task, irrelevant to the main manipulation, and was implemented for the purpose of increasing participants’ belief about the other person’s reality and simultaneous engagement in the experiment. For the same purpose, after participants responded to a stimulus, they sometimes saw a message to wait for the other person/people to respond. Participants could not read the other’s person responses during the main task; equally, the other person’s response was not relevant to their task fulfilment.

We kept the intentional and attentional relationship to the stimuli between participants and the other person constant: participants’ and others’ instructions were displayed together and participants attended to the same stimuli and answered the same question at the same time as the co-participant(s). In contrast, we varied whether they shared the same perceptual (spatial) relation to the environment. We hypothesised that, if shared experience does affect perspective taking in language use, then the mere belief of jointly looking at the same stimuli from a different versus the same spatial vantage point (while being engaged in the same task) would influence the spatial perspective that people adopt to locate objects in the jointly-attended environment. In particular, we expected participants to be more likely to produce a non-self perspective description when they were aware that their co-participant had an opposite vantage point, as the non-self description corresponded to the co-participant’s self perspective.

We should acknowledge that a lot of the evidence in favour of the task co-engagement effect comes from complementary paradigms in which parts of a task were distributed between two participants (Dolk et al., 2014). Even if individual performance success was independent,
participants were explicitly given a task to perform together, and thus shared responsibility for the overall performance. This aspect of complementarity might thus have induced a feeling of togetherness or “joint-ness” in participants (i.e., I am doing X because you are doing non-X, and we are both trying to achieve Y). It may be, hence, that it is this induced aspect of complementary, and not the belief of active co-attention, that induced people to take the other person's task into account (see however Shteynberg & Apfelbaum, 2013 for different results). If complementarity is a necessary requirement for participants’ performance to be influenced by what the co-actor is doing, then we should find that merely believing that another person is doing the same things at the same time is not sufficient to induce participants to process the co-experiencer’s perspective and shape their description of space in our study.

Moreover, unlike the previous studies that investigated the role of shared experience on perspective taking (Böckler et al., 2011), which looked at whether people used the co-participant as a reference point to process visual stimuli as measured through reaction times, we were interested in the perspective that people express through language. Most experiments that looked at the effects of shared experience and joint action on individual cognition, even those focusing on verbal responses (e.g., Philipp & Prinz, 2010), have based their findings on reaction-time measurements. It may well be that the awareness of active shared experience only affects individuals' lower-level cognitive processing but not higher-level processing that involve some form of decision planning, like those involved in utterance generation. For instance, studies that looked at the effects of another person's different focus of attention found that participants were sensitive to this information in a way that affected their performance, but this influence did not concern the actual content of their focus (the what) (e.g., Böckler & Sebanz, 2012). If so, it may be that the participants in our studies, even if affected by the other person's vantage point, would not specify the actual content of that vantage point; as a result, no effect should occur at the level of their verbal descriptions of space.

However, separate evidence suggests that the awareness of how others perceive the world (and not just that they perceive it differently), may affect the content of our experience. Firstly, research on simultaneous co-experience (non “joint action”) (e.g., Richardson et al.,
2009, Experiment 2; Boothby et al., 2014) found effects related to the content of the co-participant's experience (e.g., what the other may be paying attention to or how the other may judge a piece of chocolate). Secondly, there is some evidence that active shared experience may influence the content of one's language production even in studies that did not find such an effect at the level of reaction times. In Gambi et al. (2015) people made more naming mistakes when they believed their co-participants had to produce an incongruent response (e.g., different naming order) than a congruent response or no response at all (Experiment 1), suggesting that people might have specified the content of what the co-participant had to say. Finally, two very recent studies points to people processing the content of someone else’s perspective (i.e., how things look to them) with whom they are actively sharing a task, in a way that changes how they encode visual stimuli, even if these studies did not look at linguistic responses. In Surtees, Samson and Apperly (2016), participants judged the magnitude of numbers that looked the same (e.g., 8) or different (e.g., 6) from their perspective compared to their partner’s perspective who was seated opposite them. Participants were faster and more accurate when their perspectives were consistent (i.e., 8 vs. 6), suggesting that participants were sensitive to how numbers looked to the other. Bockler and Zwickel (2013) had pairs of participants counting male (vs. female) faces displayed on a screen located flat on a table while varying the orientation of the faces and participants’ position with respect to each other and the screen (same side vs. opposite side). They used EEG to investigate how faces at different orientations were processed, also depending on the perspective of the other person. Results showed that people considered the co-participant’s opposite perspective when processing upright (yet not inverted) faces, supporting the hypothesis that people may be responsive not only to the fact that “a co-participant sees a jointly attended scene differently but also how the other perceives it” (p. 740).

8.2 Experiment 8

8.2.1 Rationale and Predictions

Experiment 8 was designed to investigate whether the awareness of shared experience affects the perspective that people take to describe the spatial relationships between objects in the jointly-attended environment.
Participants saw photographed scenes of two objects on a table and were asked to locate one object with respect to the other (e.g., *On which side of the sugar pot is the pear?*). The objects had no clear left/right or front/back. At the beginning of the experimental session, participants were induced to believe that they were getting connected to another participant who would be doing the experiment at the same time as them (in reality a pre-set computer program). In a within-subjects blocked design, we manipulated whether the participant believed that the co-participant was sharing the same versus an opposite perspective on the scenes. We hypothesised that, if shared experience does indeed affect perspective taking in language use, then the belief of which perspective the other participant has on the jointly-attended scene should influence the spatial perspective that participants adopt to locate the objects. In particular, we expected participants to be more likely to produce a non-self description when they were aware that their co-participant had an opposite vantage point, as the non-self description corresponded to the co-participant’s self perspective.

### 8.2.2 Method

#### 8.2.2.1 Participants

A total of 288 participants were recruited via the crowdsourcing platform CrowdFlower. They were paid $1 for their participation upon completion of the task.

#### 8.2.2.2 Design

In a within-subjects between-blocks design, we manipulated whether the (fictitious) co-participant shared the same vantage point on the scenes as the participant (i.e., they were looking at the same scene photographed from the same side of the table) or, instead, had an opposite vantage point on the scenes (i.e., the other person was looking at the same scenes which were though photographed from the other side of the table) (OtherView: Same vs. Opposite). The order of the two conditions was counterbalanced across participants (Block: Block1 vs. Block2).

#### 8.2.2.3 Material

We constructed 36 experimental items, each comprising a pair of symmetrical objects (e.g., pear/sugar pot). For each item we constructed two photographed scenes, one for each
possible position of the two objects on the table (e.g., pear on the left and sugar pot on the right; sugar pot on the left and pear on the right), for a total of 72 photographed scenes. Each scene was paired with a question of the form “On which side of X is Y?” (e.g., “On which side of the sugar pot is the pear?”).

We also constructed 24 filler items, each displaying two of the objects used for the experimental items and a person, 12 males and 12 females (“person-fillers”). In half of these fillers the person was seated behind the table facing the participant, thus bearing an opposite vantage point on the objects to the participant. In the other half, the person was seated on the side of the table closer to the participant with their back to them, thus bearing the same vantage point on the objects as the participant. We also created 21 additional filler items (“neutral-fillers”), each displaying two objects, different from the ones used for the experimental items. Example experimental and filler items are shown in Figure 36.

We created 4 lists and 8 orders for each list as follows. We first divided the 72 scenes into two sets (A and B) so that each experimental item appeared only once in each set (e.g., set A contained the scene “pear on the left and sugar pot on the right”, set B contained the scene “sugar pot on the left and pear on the right”). We then divided each set into two subsets (A1 and A2 from list A; B1 and B2 from list B) and constructed four lists by counterbalancing the order of the two subsets (i.e., A1-A2; A2-A1; B1-B2; B2-B1); each subset was assigned to one experimental block in the experiment (e.g., A1 to block1 and A2 to block2 for the A1-A2 list). Each list also comprised 12 person-fillers (6 opposite and 6 same vantage point) and 18 neutral-fillers. In half the fillers, participants were asked to describe the colour of one of the object (e.g., “What colour is the [lamp]?”). In the other half, participants were asked to compare the size of the two objects (e.g., “Compared to the [grater], how big is the [bag]?”). 6 person-fillers and 9 neutral-fillers appeared in each sublist (e.g., A1 and A2). Each list also contained 14 “Please wait for the other participant to respond...” messages, 7 per sublist; these were prompts that appeared on the participant’s screen informing them that their co-participant was taking longer to respond in that trial. The duration of this message varied between 2 and 4 seconds. Each sublist started with a series of four fillers (neutral – person opposite/same (counterbalanced) – neutral) and a waiting message. Of each list we created 8 versions by randomizing the order of the experimental items, the filler items and the waiting
message durations (separately). A total of 32 lists resulted. The randomisation was constrained so that there were no more than two experimental trials in a row. For each list, four versions were assigned to the opposite OtherView first block order and the remaining four to the same OtherView first block order; in this way, all items appeared once in each block under both conditions. Participants were randomly assigned to one of the 32 lists.

![Example photograph scenes used as stimuli in the experiments: experimental item (a, b); female person-filler opposite vantage point (c); male person-filler same vantage point (d); neutral filler (e).](image)

8.2.2.4 Procedure

On the CrowdFlower platform, participants read the information introducing the study and gave their informed consent to take part in the experiment (see Appendix R). They were told that they would do the experiment at the same time as another person; for this reason, at the beginning of the experiment we would try to locate and get them connected to another CrowdFlower worker, and that this might take up to a few seconds. Concerning the actual task, they were simply instructed that they would be seeing a series of photos and be asked a question about their content. They were then asked to follow a link to an external website, hosted on the university's server, where the experiment was run. By clicking on the link they
were randomly assigned to one of the 32 experimental lists. In the initial screens participants provided information about their gender, age and the first language learned as a child. The information about their gender was used to assign them to a same-gender version of the list (e.g., for male participants: the other participant was a male and a man was in the person-filler). Next, participants saw a “connection screen” (cf. Duran & Dale, 2012) where participants waited for the software to locate and connect them to another participant. The screen lasted 15 seconds, after which a prompt appeared on the screen informing participants that another participant had been located and they were now been connected together.

In order to enhance the illusion of connectivity participants were then asked to introduce themselves to the other person and provide them with the code that they could see on their screen, ostensibly to verify that the connection worked (“Connection Verification screens”). On the next screen, they received the same information from the other participant (in reality, pre-set information) and typed in the received code. The other participant was Andrew if male and Lauren if female. To increase the chance of participants believing in the sham connection, they then engaged in the first part of an intuition task; at the beginning of the two blocks participants described for the other person three things that they liked or liked doing, and three things that they disliked or disliked doing. They would receive the same information from the other person. They were told that at the end of the experiment they would be asked to guess the identity of the other person (e.g., age, location, profession) based on the information received. The information given by the fictitious participant was the same across conditions and gender (i.e., the co-participant liked trying new cafes, cycling to work, and checking the news first thing in the morning; they disliked crowded places, waiting in queues and spicy food), as well as their guess about the (real) participant's identity (i.e., “So, I think you are in your late 20s, from Southwest United States, perhaps Texas, and are working as a salesperson”).

Participants received detailed instructions at the beginning of each block. In both conditions, they were told that they and [Andrew/Lauren] would be looking at a series of photographed scenes of two objects on a table. Each time, they would be looking simultaneously at the same scene and be asked the same question about the objects (e.g., how big they were, or where they were). In the same OtherView condition, participants were told that the
photographs that the two of them would see were taken from the same side of the table. On the contrary, in the opposite OtherView condition, participants were told that despite looking simultaneously at the same scene, the photographs that the two of them would see were taken from the opposite sides of the table. Examples of photos from both or the one vantage point were shown. After each block instructions, participants were presented with a validation question that asked them whether the photographs that they and [Andrew/Lauren] would see were taken from the same side vs. opposite sides of the table.

At the end of the second experimental block, we carried out two manipulation checks. As implicit manipulation check, we used participants' guess about the identity of the other person. We assumed that people who suspected having done the experiment with a computer partner, rather than a real person, would spontaneously state their suspicion here. We then carried out an explicit manipulation check and asked participants whether they believed the other participant was a real person or a computer.

8.2.3 Analysis

8.2.3.1 Coding and Exclusion Criteria

10 participants dropped out of the online experiment before this reached an end and they were not retained for analysis. Hence, the initial sample consisted of 278 participants. Of these, 2 participants had already taken part in one of our previous experiments, 29 did not have English as their first language, 2 did not comply with the instructions (e.g., they typed in random strings of characters), and 21 failed either or both validation questions assessing whether they had read and understood the block instructions. Their data were not retained for analysis. Of the remaining 8064 descriptions (from 224 participants), 38 (0.47%) expressed no perspective (e.g., “on the front”, “the other side of the table”) and were removed from the dataset. 36 of these 38 descriptions came from the same participant who never described object locations in terms of left/right.

The remaining 8028 descriptions (from 223 participants) located the target object as being on the left or the right of the referent object. Descriptions were categorised as being from a self perspective if they were from the participant’s own perspective and non-self otherwise. 34 descriptions (0.42%) explicitly accounted for both possible perspectives (e.g., “My side: left,
Lauren’s side: right”). They were produced by two participants: one participant accounted for both perspectives throughout the whole Opposite OtherView block, and a second participant accounted for both perspectives throughout the whole Opposite OtherView block with the exception of two trials. As this type of responses only accounted for 0.43% of observations, the responses were excluded from analyses.

13 participants (4.7%) did not pass our implicit manipulation check; when asked to make a guess about the identity of the other participant, they either spontaneously stated that the other participant was in fact a computer (n = 7) or failed to make any guess (n = 6). We excluded them from analysis. Our final dataset was composed by 7449 descriptions from 208 participants.

8.2.3.2 Model and data exploration

We used a two-part or hurdle mixed model to analyse which factors might induce participants to adopt a non-self perspective (Min & Agresti, 2005; Zuur, Saveliev, & Ieno, 2012). We chose this approach due to the many participants in our sample who never produced a non-self description. As this is not the usual approach in spatial-language perspective research, we will briefly discuss some data exploration issues. Figure 37 shows the frequency plots for the data. The x-axis displays the observed number of non-self descriptions per OtherView condition (i.e., maximum possible value = 18) and the y-axis displays how often each value was observed (i.e., how many participants produced that number of non-self descriptions).
The data contain 47% zeros in the opposite OtherView and 48% zeros in the same OtherView (i.e., participants who produced no non-self descriptions in that condition). This is data for which a generalized linear model (GLM and GLMM) which assumes a Poisson or Binomial distribution may have lack of fit due to disproportionately many zeros, producing biased parameter estimates (Min & Agresti, 2005; Zuur et al., 2012). Moreover, the hurdle approach allows for the possibility that the mechanisms that determine whether a person ever switches to a non-self perspective at all may be different from the mechanisms that determine how many non-self descriptions are produced by the people who do in fact switch (cf. Ridout, Demétrio, & Hinde, 1998). This allows investigating interesting theoretical distinctions. For instance, there might be people that are naturally less prone to take perspectives other than the self or are just less responsive to social factors (cf. Wardlow, 2013; we return to this possibility in the General Discussion).

In the first part of the hurdle mixed analysis, a binomial distribution with a logit link function was used to model the probability that a participant switched at least once to a non-self perspective ($y = 1$) or never switched ($y = 0$) in relation to the experimental OtherView-by-Block conditions. In the second part, participants’ positive count responses (zeros were

Figure 37 Frequencies of non-self descriptions produced by participants in each OtherView condition. Participants were tested in both conditions.
dropped) were modelled using a truncated Negative Binomial distribution (log link function) to investigate whether OtherView (opposite vs. same) and the order with which the conditions were presented affected the frequency with which participants produced non-self descriptions. The choice of a (truncated) Negative Binomial (vs. Poisson) distribution was done to account for the over-dispersion caused by the fact that a participant's responses were not independent events; a participant producing a non-self response may spark off other non-self responses on later trials ('contagious events'). Thus, the two-part model approach allowed us to investigate which predictors drove the probability to ever switch to a non-self perspective and, if a speaker did switch at least once, which predictors drove the number of non self-descriptions produced (i.e., the extent to which they adopted a non-self perspective). Figure 38 illustrates the two steps in the analysis.

Figure 38 In the first part of the analysis (left panel), participants are assigned a value of 1 if they have produced a non-self description at least once in that condition, and 0 otherwise. A binomial GLMM with a logit link function was used to model the data, with a by-participant random intercept to account for the within-subjects data dependencies (red arrow in the panel). In the second part of the analysis (right panel), all zero-counts are dropped and a truncated Negative Binomial GLMM with a log link function and by-subject random intercept is applied to the positive counts only. This means that if, for instance, a participant produced at least one non-self description (i.e., “switched”) in the opposite OtherView condition, but no non-self descriptions in the same OtherView condition (e.g., the pink participant in the middle), in this second part of the analysis the participant's responses in the same OtherView condition are dropped. A (truncated) Negative Binomial was used instead of a (truncated) Poisson distribution due to the over-dispersion caused by a participant's responses being not independent events ('contagious event').
Predictors were numerically coded and centred to reduce collinearity and obtain ANOVA-like interpretation of main effects (this equals to deviation coding $0.5/-0.5$ if the design is balanced), unless otherwise specified. The interpretation of the coefficients from the binomial GLM(M) is as in previous chapters. Coefficient parameters from the truncated Negative Binomial regression represent the expected difference in log counts between the two levels of the categorical predictor. Exponentiated coefficients represent incident rate ratios (IRR) where 'incident', in our case, is the production of a non-self description. Similar to log odds for logit regression, a significant positive coefficient indicates a higher rate (to produce a non-self descriptions) for the predictor level coded positively (e.g., $0.5$) versus the level coded negatively, the opposite for a significant negative coefficient. The significance of the predictors was evaluated using Wald's Z statistic, which assesses whether the regression coefficients are significantly different from zero (given the estimated S.E.). All analyses were run in R (version 3.1.3, R Code Team, 2014). We used the \{glmmADMB\} package (version 0.8.3.3) (Skag, Fournier, Bolker, Magnusson, & Nielsen, 2015) when the design involved a within-subject manipulation, and the \{pscl\} package (version 1.4.9) (Zeileis, Kleiber, & Jackman, 2008) when all manipulations were between-subjects.

In Appendix S we provide further justification for the two-stage hurdle approach and for treating our data (originally generated by a binary process) as count.

8.2.4 Results

Data and results are reported following the two stages of the hurdle analysis\(^{46}\).

Hurdle Model Part-I: Binomial GLMM for whether participants switched perspectives at least once

Table 25 and Figure 39 show frequency and proportion of participants who switched to a non-self perspective at least once in relation to OtherView and Block.

\(^{46}\)Visual inspection of mean rate of non-self descriptions by items revealed an outlier. Excluding it didn't change the pattern of results in either part of the analysis.
### Table 25

Frequency of participants who switched (never switched) in Experiment 8 according to OtherView and Block.

<table>
<thead>
<tr>
<th>OtherView</th>
<th>Block 1</th>
<th>Block 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opposite</td>
<td>58 (38)</td>
<td>52 (62)</td>
</tr>
<tr>
<td>Same</td>
<td>64 (50)</td>
<td>43 (53)</td>
</tr>
</tbody>
</table>

We fitted a GLMM with a logit link-function and a by-subject random intercept. OtherView (Opposite = .5 / Same = -.5) and Block (Block1 = .5 / Block2 = -.5) were included as fixed-effects predictors. Participants were no more likely to ever switch to a non-self perspective in the Opposite than the Same OtherView condition (no effect of OtherView; \( \beta = 0.12, SE = 0.21, \) odds-ratio = 1.13, 95% CI [0.75; 1.70], \( p = .572 \)). However, participants were more likely to ever switch in Block 1 compared to Block 2 (significant effect of Block; \( \beta = 0.60, SE = 0.22, \) odds-ratio = 1.82, 95% CI [1.19; 2.76], \( p = .006 \)). This tendency was not mediated by OtherView (no significant OtherView by Block interaction; \( \beta = 0.17, SE = 0.46, \) ratio of odds-ratio = 1.18, 95% CI [0.48; 2.94], \( p = .714 \)).

**Figure 39** Observed proportion of participants who switched to a non-self perspective at least once per experimental condition in Experiment 8. Error bars denote non-parametrically bootstrapped (BCa) 95% confidence intervals.
Hurdle Model Part-II: Truncated negative binomial GLMM for the positive count responses

For the second part, we modelled the number of non-self descriptions produced by participants who switched at least once in that condition using a truncated negative binomial distribution ($\log$). Observed mean rates of non-self descriptions are reported in Table 26 and Figure 40. We included OtherView (Opposite = .49 / Same = -.51) and Block (Block1 = .44 / Block2 = -.56) as fixed-effects predictors, and a by-subjects random intercept.

Table 26
Mean rate (variance) of non-self descriptions in Experiment 8 according to OtherView and Block (positive count responses only)

<table>
<thead>
<tr>
<th>OtherView</th>
<th>Block 1</th>
<th>Block 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opposite</td>
<td>3.40 (17.05)</td>
<td>2.42 (11.78)</td>
</tr>
<tr>
<td>Same</td>
<td>1.89 (2.07)</td>
<td>3.51 (20.35)</td>
</tr>
</tbody>
</table>

Figure 40 Observed mean rates of non-self descriptions in relation to OtherView and Block in Experiment 8. Error bars denote non-parametrically bootstrapped (BCa) 95% confidence intervals on subject-wise condition mean rates.
No main effect was significant (OtherView β = 0.23, SE = 0.17, IRR = 1.26, 95% CI [0.91; 1.76], p = 0.166; Block β = -0.06, SE = 0.17, IRR = 0.94, 95% CI [0.67; 1.31], p = 0.708). The model showed a marginally significant OtherView by Block crossover interaction (β = 1.09, SE = 0.59, ratio of IRR = 2.96, 95% CI [0.94; 9.35], p = 0.064), suggesting that the order with which participants completed the two conditions modulated the results.

Even if the interaction was only marginally significant, we decided to investigate it further. Exploratory post-hoc analyses comparing the number of non-self descriptions produced in the Same vs. Opposite OtherView separately for each block showed a reliable order effects; participants produced significantly more non-self descriptions in the Opposite compared to the Same OtherView condition in Block 1 (β = 1.17, SE = 0.37, IRR = 3.23, 95% CI [1.57; 6.65], p = .007), whilst no significant difference was found in Block 2 (β = -0.68, SE = .42, IRR = 0.51, 95% CI [0.22; 1.14], p = .200). Results seem to suggest a possible carryover effect; participants who were exposed to the Same OtherView condition after the Opposite condition may have been unable to shed the alternative vantage point in Block 2 even if this was no longer the other person's current vantage point. This interpretation is supported by the fact that the number of non-self descriptions did not depend on block order for the Opposite OtherView (β = 0.62, SE = 0.38, IRR = 1.86, 95% CI [0.88; 3.95], p = .200), while it did for the Same OtherView (β = 1.23, SE = 0.40, IRR = 3.42, 95% CI [1.56; 7.46], p = .007). This suggests that while non-self perspective taking occurred in the Opposite OtherView condition irrespective of order, participants were influenced by the opposite vantage point in the Same OtherView condition only when they had first participated in the Opposite condition.

We then looked at whether participants' reported belief about their co-participant's identity (i.e., real person vs. a computer) modulated the effects. Overall, 33% of participants (69/208) reported that the other was a real person when they were explicitly asked about it. Table 27 and Figure 41 show frequency and proportion of those who ever switched versus never switch according to OtherView and whether they believed the other participant was a real person or a computer.

---

For the whole remaining of this section, all reported p-values were adjusted using Benjamini and Hochberg's False Discovery Rate (B&H's FDR).
Table 28 and Figure 42 report mean rate of non-self descriptions among those who switched at least once according to the same variables. We restricted our analysis to block 1 only, in order to isolate the effect of first exposure to each OtherView condition, and re-ran the two-stage hurdle model including co-participant Reported_Identity (real person = .66 / computer = -.34) as a fixed effect predictor together with its interaction with OtherView (Opposite = .54 / Same = -.46).

The logistic regression for the binary data found no statistically significant effect of either OtherView, Reported_Identity, or of their interaction (all ps > .3): no variable affected the probability of participants to ever switch to a non-self perspective. The truncated negative binomial regression for the positive count data confirmed a significant main effect of OtherView (β = 1.14, SE = 0.37, IRR = 3.14, 95% CI [1.51; 6.49], p = 0.007): the participants who switched produced more non-self descriptions if their co-participant held an opposite perspective compared to them (vs. the same perspective). The main effect of Reported_Identity was not significant (β = 0.35, SE = 0.41, IRR = 1.42, 95% CI [0.64; 3.15], p = .488). The effect of OtherView was not modulated by participants’ belief about their co-participant's identity as revealed by the non-significant interaction (β = -0.26, SE = 0.81, ratio of IRR = 0.77, 95% CI [0.16; 3.73], p = .750).
Table 27

Frequency of participants who switched (never switched) in block 1 of Experiment 8 according to OtherView and belief about the co-participant's Reported_Identity.

<table>
<thead>
<tr>
<th>Reported_Identity</th>
<th>OtherView</th>
<th>real person</th>
<th>computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opposite</td>
<td>14 (21)</td>
<td>24 (37)</td>
<td></td>
</tr>
<tr>
<td>Same</td>
<td>20 (16)</td>
<td>30 (48)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 41 Observed proportion of participants who switched to a non-self perspective at least once in relation to OtherView and Reported_Identity in Block 1 of Experiment 8. Error bars denote non-parametrically bootstrapped (BCa) 95% confidence intervals.

Table 28

Mean rate (SE) of non-self descriptions in block 1 of Experiment 8 according to OtherView and participants’ reported belief about their co-participant's identity (positive counts only)

<table>
<thead>
<tr>
<th>Reported_Identity</th>
<th>OtherView</th>
<th>real person</th>
<th>computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>opposite</td>
<td>3.67 (13.83)</td>
<td>3.24 (19.24)</td>
<td></td>
</tr>
<tr>
<td>same</td>
<td>2.19 (1.90)</td>
<td>1.79 (2.13)</td>
<td></td>
</tr>
</tbody>
</table>
8.2.5 Discussion

Knowing that another person was attending to the same scene and carrying out the same spatial task at the same time as them affected the perspective that people adopted to locate objects within that scene. Specifically, how often participants opted for a non-self perspective was influenced by the other person's vantage point; participants produced more non-self descriptions if the other person was thought to be seeing the scene photographed from the opposite side of the table, presumably as a result of adopting the other person's perspective. However, the effect was limited to block 1, i.e., to people's first experience in the experimental setting. Once participants were aware of the other possible alternative perspective on the scene, they retained and continued being influenced by it even in the subsequent block despite the other participant now sharing the same perspective. This is in line with Atmaca et al. (2011), who found a similar carry-over effect in a joint Flanker task (experiment 3). In their study, they found that once participants entered in a “joint mood” in block 1 (i.e., they believed they were doing the task together with another person), they kept

![Figure 42 Observed mean rates of non-self descriptions in relation to OtherView and Reported Identity in block 1 of Experiment 8. Error bars denote non-parametrically bootstrapped (BCa) 95% confidence intervals on subject-wise condition mean rates.](chart.png)
being influenced by the person's task to a similar extent also in block 2 even if they were now told that they were doing the experiment alone.

The effect was not modulated by whether participants reported that the co-participant was a real person or not. This is important, as it might have undermined the role for shared experience in our results. Shared experience requires in fact that the other person is believed to be real. Thus, we cannot rule out that the effect found might simply be caused by the alternative perspective made salient by the instructions (and so it is independent of the co-participant's existence). Finally, we did not find that knowing the other person's vantage point on the jointly-attended scenes predicted whether people would switch to a non-self perspective at least once (see General Discussion).

8.3 EXPERIMENT 9

8.3.1 Rationale

In Experiment 9, we tried to better understand the effect found in Block 1 of Experiment 8 by using a between-subject design that ruled out possible carry-over effects from Block 1 to Block 2. We also investigated whether the effect could be extended to a group of other people. That is, we were interested to see whether believing that several another people were jointly attending to the same environment yet from an opposite perspective could also affect the perspective with which a person would locate an object in space. Furthermore, we were interested in a possible “crowd effect”: would the belief that a crowd of people (compared to a single another person) is simultaneously attending to the same environment enhance the chance with which individuals take those people’s perspective into account to describe space? Existing findings in attention research, for instance, have found that when social information is available in crowds, people’s reactions are amplified (e.g., Milner, Bickman, & Berkowitz, 1969). For example, a crowd’s direction of looking is more effective than an individual’s gaze for directing one’s attention (Gallup et al., 2012; Sweeny & Whitney, 2014).

We expected to replicate the effect found in Block 1 of Experiment 8 for the Solo Other condition: if the participant's and the co-participant's vantage points differed (Opposite OtherView), the participant should produce more non-self descriptions compared to when
their vantage points matched (Same OtherView). We expected to find the effect of OtherView also when “the co-participant” was a group of co-participants. In addition, we hypothesised a “crowd effect”: the effect of OtherView on the perspective adopted by participants should be enhanced if participants believed that a group of co-participants (vs. a single person) was having an opposite perspective to theirs. Knowing that a group of people are sharing an opposite perspective to the one’s own should make that opposite perspective more salient (and so more able to influence the perspective one chooses to adopt to locate objects).

8.3.2 Method

8.3.2.1 Participants

A total of 384 participants were recruited via the crowdsourcing platform CrowdFlower. They were paid $1 for their participation upon completion of the task.

8.3.2.2 Design

In a between-subjects design, we manipulated whether the participant believed they did the experiment at the same time as one other person or five another people (Other: Solo vs. Crowd), and whether the (fictitious) co-participant(s) had the same versus the opposite vantage point on the scenes (OtherView: Same vs. Opposite). We included the participant’s reported belief about the co-participant(s)’s identity as a co-variate (Reported_Identity: real person vs. computer). We chose a crowd size of five since previous studies on crowd effect on gaze-direction (Milner et al., 1969) found that the effect increases up to that size and then saturates.

8.3.2.3 Material

As in Experiment 8 with the following differences: 128 lists were created (four randomised versions of 32 original lists). Participants were randomly assigned to one of these lists (three per list).
8.3.2.4 Procedure

This was as in Experiment 8 but with the following differences: Participants were exposed to the same experimental condition in block 1 and block 2. When the intuition test was introduced, participants were asked to provide the information that they had to guess about each other (i.e., age, location, profession), ostensibly so that their guesses could be assessed. The experiment was only available for participants to take part on a daily two-hour timeslot (10-12:00 Eastern Daylight Time) to make it more believable that 6 people could be recruited at the same time within a matter of seconds.

In the crowd Other condition, participants were connected to another five people (either 3 male and 2 female or 2 male and 3 female), and received names and codes from all five of them. Participants then verified the connection with only one of the five other people, chosen at random (male or female, counterbalanced). This was the person they would be asked to guess the identity of and to whom they were providing and from whom they were receiving information for the intuition test. The waiting messages emphasised that they had to wait for all the other people to respond. Overall, half of the recruited male participants were assigned to a same gender co-participant, and the remaining half to an opposite gender co-participant (similarly for female participants). Gender affected the gender of the communicating co-participant and of the persons in the filler trials.

8.3.3 Analysis

8.3.3.1 Coding and Exclusion Criteria

64 participants had already taken part in our previous experiments, 45 did not have English as their first language, and 51 failed either validation questions assessing whether they had read and understood the block instructions. Their data were excluded for analysis. We further removed from the dataset 76 responses (0.95%) due to problems with photo uploading (6) or because participants typed in nonsense strings (1 participant did so for the whole experiment), and 84 further descriptions (1.04%) which expressed no perspective (e.g., “on the front”; of these, 36 came from a single participant who never described object locations in terms of left/right).
The remaining 7904 descriptions (from 222 participants) located the target object as being on the left or the right of the referent object. Descriptions were categorised as being from a ‘self’ perspective if they were from the participant’s own perspective and ‘non-self” otherwise. 62 descriptions explicitly accounted for both possible perspectives. They were produced by two participants, one of which did it for all the experimental trials. As they only accounted for 0.78% of observations, they were excluded from analyses. Of the remaining 7842 descriptions from 221 participants, 55 (0.7%) expressed explicitly a self-perspective (e.g., “my left”) and 1 explicitly a non-self perspective (i.e., “your right”). They were retained in the analysis.

13 participants (3.4%) did not pass our implicit manipulation check; when asked to make a guess about the identity of the other participant, they either spontaneously stated that the other participant was in fact a computer (n = 4) or failed to make any guess (n = 9).

Our final dataset was composed by 7405 descriptions from 208 participants (49 in the Crowd / Same, 54 in the Crowd / Opposite, 54 in the Solo / Same and 51 in the Solo / Opposite).

8.3.3.2 Model

As for Experiment 8 with the following differences: We analysed the Other Solo and Crowd conditions separately, with the former being a direct replication of Experiment 8 / Block 1. In both parts of the hurdle model, we included OtherView, partner Reported_Identity and their interaction as predictors. Both variables were numerically coded and centred.

Empirical density plots are reported in Appendix T together with a comparison of the model predictions from the hurdle model (hereafter reported) and a Negative Binomial GLM.

8.3.4 Results

Table 29 and Figure 43 shows frequency and proportion of participants who switched to a non-self perspective at least once in relation to Other, OtherView and Reported_Identity.
Table 30 and Figure 44 report observed mean rates of non-self descriptions (positive counts only). We first analysed the solo and the crowd Other conditions separately as we were interested in whether we could replicate the opposite vs. same OtherView effect that we found in Block 1 of Experiment 8, and whether the same effect could also be found in the crowd condition. We then compared the effects between the crowd and the solo conditions to investigate whether the effect was enhanced by the crowd (“crowd effect”).

Table 29
Frequency of participants who switched (never switched) at least once to a non-self perspective in Experiment 9.

<table>
<thead>
<tr>
<th>OtherView</th>
<th>Solo Other</th>
<th>Crowd Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>computer</td>
<td>real person</td>
</tr>
<tr>
<td>Opposite</td>
<td>26 (7)</td>
<td>14 (4)</td>
</tr>
<tr>
<td>Same</td>
<td>33 (5)</td>
<td>12 (4)</td>
</tr>
</tbody>
</table>

Figure 43 Observed proportions of participants who ever switched to a non-self perspective in relation to Other, OtherView and Reported Identity in Experiment 9. Error bars denote non-parametrically bootstrapped (BCa) 95% confidence intervals.
Table 30

Mean rate (variance) of non-self descriptions in Experiment 9 according to Other, OtherView and participants' reported belief about their co-participant's identity (positive counts only).

<table>
<thead>
<tr>
<th>OtherView</th>
<th>Solo Other</th>
<th>Crowd Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>computer</td>
<td>real person</td>
</tr>
<tr>
<td>Opposite</td>
<td>4.23 (20.43)</td>
<td>4.21 (48.49)</td>
</tr>
<tr>
<td>Same</td>
<td>4.06 (22.56)</td>
<td>1.50 (0.45)</td>
</tr>
</tbody>
</table>

Figure 44 Observed mean rates of non-self descriptions in relation to Other, OtherView and Reported_Identity in Experiment 9 (positive counts only). Error bars denote non-parametrically bootstrapped (BCa) 95% confidence intervals on subject-wise condition mean rates.
8.3.4.1 Direct replication: Solo Other

Part-I: Binomial GLM (logit) for whether participants ever switched to a non-self perspective

The binomial regression with OtherView (Opposite = .51/ Same = -.49) and co-participant's Reported_Identity (real person = .67 / computer = -.33) as predictors found no statistically significant association between either predictor and whether participants ever switched to a non-self perspective. Participants were no more likely to ever use a non-self perspective if the co-participant had an opposite (vs. the same) view on the scene ($\beta = -0.41$, SE = 0.50, odds-ratio = 0.66, 95% CI [0.25; 1.78], $p = .415$), nor if they reported to believe that the other was a real person vs. a computer ($\beta = -0.54$, SE = 0.51, odds-ratio = .58, 95% CI [0.22; 1.57], $p = .285$). The OtherView by Reported_Identity interaction was also not significant ($\beta = 0.51$, SE = 1.01, ratio of odds-ratio = 1.66, 95% CI [0.23; 11.96], $p = .616$).

Part-II: Truncated negative binomial GLM (log) for the positive count data only

For the second part, we modelled the subset of data from participants who produced a non-self perspective at least once, using the positive count of their non-self responses as outcome variable.

The truncated negative binomial regression revealed a marginally significant effect of OtherView ($\beta = .71$, SE = 0.41, IRR = 2.03, 95% CI [0.92; 4.50], $p = .080$), a significant main effect of co-participant's Reported_Identity ($\beta = -1.02$, SE = 0.45, IRR = 0.36, 95% CI [0.15; 0.87], $p = .024$) and a significant OtherView by Reported_Identity interaction ($\beta = 2.08$, SE = 0.90, IRR = 8.0, 95% CI [1.36; 47.12], $p = .021$). In line with our hypothesis, participants produced more non-self descriptions when they knew their co-participant was having an opposite vantage point on the jointly-attended scenes, thus replicating our findings in Block 1 of Experiment 8. However, unlike in Experiment 8, the effect was limited to those participants who reported believing their co-participant was a real person (vs. a computer). Post-hoc analyses comparing the Same vs. Opposite OtherView in the two groups separately (i.e., those who believed the co-participant was a real person vs. not) confirmed

---

48 B&H's FDR adjusted p-values are reported.
this modulatory effect of Reported_Identity: participants who reported believing in the reality of their co-participants produced more non-self descriptions if their co-participant held an opposite vantage point (β = 2.08, SE = 0.77, IRR = 8.0, p = .01), whilst OtherView did not affect the spatial perspective of those who did not believe in the reality of their co-participant (p > .5).

8.3.4.2 Crowd Other

Part-I: Binomial GLM for whether participants ever switched to a non-self perspective

The binomial regression with OtherView (Opposite = .48/ Same = -.52) and co-participant's Reported_Identity (real person = .65 / computer = -.35) as predictors found no statistically significant association between either predictor and whether participants ever switched to a non-self perspective. Participants were no more likely to ever use a non-self perspective if the crowd of co-participants had an opposite (vs. the same) view on the scene (OtherView β = 0.48, SE = 0.52, odds-ratio = 1.62, 95% CI [0.58; 4.52], p = .356), nor if they reported believing that others were real people vs. a computer (β = .002, SE = 0.56, odds-ratio = 1.00, 95% CI [0.33; 3.01], p = .998). The OtherView by Reported_Identity interaction was also not significant (β = 0.98, SE = 1.12, ratio of odds-ratio = 2.67, 95% CI [0.30; 23.75], p = .379).

Part-II: Truncated negative binomial GLM for the likelihood to produce a non-self perspective (positive count only)

The truncated negative binomial regression revealed no significant main effect or interaction. Participants did not produced more non-self descriptions when the crowd of co-participants held an opposite (vs. the same) perspective on the scene (OtherView β = -0.47, SE = 0.44, IRR = 0.63, 95% CI [0.27; 1.47], p = .282) or when they believed (vs. did not believe) in the reality of their co-participants (Reported_Identity β = -.22, SE = 0.44, IRR = 1.00, 95% CI [0.33; 1.96], p = .282). The OtherView by reported_Identity interaction was also not significant (β = -0.48, SE = 0.91, IRR = 2.67, 95% CI [0.10; 3.74], p = .604).
8.3.4.3 The “crowd effect”

We defined the crowd effect as the enhanced effect of the other(s)’s opposite perspective on participants’ likelihood to produce a non-self description in the crowd compared to the solo group. To test it, we built a hurdle model comparing the effect of the co-participant’s perspective between the solo and crowd groups. The model included Other (Solo = .5; Crowd = -.5), OtherView (Opposite = .5; Same = -.5), Reported_Identity (real person = .66; computer = -.34) and all their two-way and three-way interactions. The crowd effect would predict a significant negative coefficient for the OtherView by Other interaction.

Part-I: Binomial GLM for whether participants ever switched to a non-self perspective

There was no significant main effect (ps > .4). Contrary to our hypothesis of a crowd effect, the OtherView by Other interaction was not significant (β = -0.88, SE = 0.72, ratio of odds-ratio = 0.42, 95% CI [0.10; 1.71], p = .224). All other two-way interactions (ps > .4) and the three-way interaction (p > .5) were also not significant.

Part-II: Truncated negative binomial GLM for the likelihood to produce a non-self perspective (positive count only)

No main effect was significant (all ps > .05). There was a significant OtherView by Other interaction (β = 1.19, SE = 0.59, ratio of IRR = 3.29, 95% CI [1.03; 10.51], p = 0.045), but the effect went in the opposite direction to that predicted by the crowd effect hypothesis: the co-participant’s opposite (vs. same) perspective increased the number of non-self descriptions in the Solo condition (i.e., when there was only one co-participant) but not in the Crowd condition (i.e., when there were many). No other two-way interaction was significant (ps > .2). There was a significant three-way interaction (β = 2.58, SE = 1.28, ratio of IRR = 13.18, 95% CI [1.07; 161.89], p = 0.044): the effect of OtherView on the number of non-self descriptions in believing participants (compared to those not believing in the reality of the co-participant) was present in the Solo but not in the Crowd condition.
8.3.5 Discussion

In the solo condition of Experiment 9 we replicated the shared-experience effect found in Block 1 of Experiment 8; people produced more spatial descriptions from a non-self perspective if they believed the other person had an opposite vantage point on the jointly-attended scene. However, in contrast to Experiment 8, the effect only occurred for those participants who, when explicitly asked, reported believing that the other participant was a real (vs. simulated) person. We did not find any effect of co-participants’ perspective in the crowd condition.

Contrary to our predictions, we did not find a “crowd effect”, understood as an enhanced shared-experience effect induced by the belief that a group of people (compared to only one person) was having an opposite perspective to one's own.

8.4 General Discussion

In two experiments, participants described the location of objects in photographed scenes after being told that another person (Experiment 8) or several other people (Experiment 9) were carrying out the experiment at the same time as them. We found that participants produced non-self perspective descriptions to different extents depending on whether they assumed the other was sharing their vantage point (i.e., fewer non-self descriptions) or not sharing their vantage point (i.e., more non-self descriptions) on the jointly-attended stimuli. Thus, our results suggest that the belief of sharing the experience with someone else made speakers responsive to the other person's vantage point, even if the other person’s presence could only be assumed within the context.

Were the speakers in our task more likely to engage in self perspective taking (producing fewer non-self descriptions when the other shared their vantage point) or more likely to engage in non-self perspective taking (producing more non-self descriptions when the other had an opposite vantage point) as a consequence of our shared-experience manipulation? If we compare the observed mean rate of non-self descriptions across the two experiments, it looks clear that “believers” (i.e., those who reported believing in the reality of the co-participant(s)) in the same OtherView condition (Block1 / Experiment 8; Solo Other / Experiment 9) produced fewer non-self descriptions than in any other condition, and were
also more consistent in doing this, as shown by the much smaller variance. This trend suggests that the effect was driven by participants producing more self descriptions when they believed their partner held their same vantage point on the jointly-attended scenes. Thus, shared experience seems to have reinforced a spatial self perspective.

Our results differ from Böckler and colleagues (2011) findings that when the other person’s spatial relation to the objects was highlighted by joint attention, people were more likely to switch to a non-self perspective. Even if the two studies are not directly comparable (e.g., different paradigms and response type, lack/presence of physical co-presence and of a same perspective condition), the reason why we did not find effects in terms of increased non-self perspective taking could be that in our task, the position of the other person wouldn't facilitate performance, as it did in their study. In contrast, our results are in line with Surtees et al. (2016), who found that participants made fewer mistakes in assessing the magnitude of numbers when these looked the same to both them and their partner (e.g., 8) compared to numbers that looked different from their partner’s vantage point (e.g., 6). As in our task, but unlike in Böckler et al.’s (2011) task, performance in Surtees et al.’s (2016) task would not benefit from processing the other person’s perspective.

How do these results relate to our findings from Experiment 5-7? In Experiment 7, the experiment whose conditions are directly comparable, we found that the vantage point of the other person modulated perspective taking both when it differed from the participant’s (i.e., more non-self descriptions when the person was located behind the table, facing towards the objects and the participant) as well as when it was the same as the participant's (i.e., fewer non-self descriptions when the person was located in front of the table, facing towards the objects but away from the participant). Why this difference in results? It could be that, in Experiment 7, the physical presence of the other person made people aware that they had a clearly different (vs. same) spatial relation towards the objects compared to him/her; a cue that our participants in Experiment 8 and 9 could not rely on. It must also be noted, however, that other studies into the effects of shared experience have found effects that go in the direction of an enhanced self perceptual perspective, with people amplifying what they themselves felt. For instance “good” chocolate was rated as better and “bad” chocolate as
worse if people believed they co-experienced the chocolate with someone else (Boothby, et al., 2014).

Another element for discussion is that while in Experiment 8 the effect was independent of people’s reported belief about the identity of the other person (real vs. simulated), in Experiment 9, the effect was only found for those who reported believing that the other person was real (the “believers”). One possibility is that the effect of shared-experience on the “non-believers” in Experiment 8 was due to people being primed by the instructions. Based on the current data, we cannot rule out this hypothesis; nor can we dismiss that this explanation also underlies the results in Experiment 9 and for the “believers” in Experiment 8. Future studies should include a non-social priming condition in which participants are told about the alternative perspective but then engage individually in the task. Note that joint action studies that have included similar control conditions did not find any effect of non-social priming on participants' performance (e.g., Bockler et al., 2011; 2012).

This difference in results between our two experiments remains, however, problematic for our investigation. Conclusions about any shared-experience effects (or lack thereof) can be made only if participants are shown to have bought into the story of the reality of the co-participant(s). It is difficult to say whether people's answer to the explicit question about the reality of the co-participant(s) really reflected what they were thinking during the experiment, or is instead the result of participants being prompted to doubt by the question at the end. Very few people spontaneously stated that they thought the other was not real in our implicit manipulation check.

Interestingly, our shared-experience manipulation only affected how often people engaged in non-self perspective taking if they did (Hurdle model – Part II) but not whether they ever used a non-self perspective at all (Hurdle model – Part I). A speculative explanation for this pattern of results is that the manipulation was not strong enough to affect the “low-switchers”, that is, those people who are naturally less likely to take other people’s perspectives and mental states into account. That people do have different predispositions in this regards has been confirmed by a number of recent studies. Wardlow (2013), for instance, found that speakers' tendency to take the partner's perspective in a referential communication task could be predicted by their working memory and executive control abilities (see also
Similar modulatory effects have been found for individuals' spatial (Schober, 2005; 2009) and social skills (Shelton, Clements-Stephens, Lam, Pak, & Murray, 2012). Our manipulation might have thus only affected those who are able to effectively recognize and imagine alternative perspectives; as a result, the observed significant difference only concerned the extent of non-self perspective taking in this group of people, and not whether people ever switched (i.e., number of high vs. low switchers). However this explanation does not account for the fact that, in Experiment 8, the same participants were more likely to ever switch in Block 1 compared to Block 2 (within-subject comparison). Follow-up studies should take account of these underlying individual differences. An interesting aspect of the Hurdle models is that they do allow us to investigate whether individual differences modulate both the tendency to switch and the extent to which people switch, and if the same factors affect both aspects to a similar degree.

Related to the point just raised about speakers’ individual differences, it is interesting to note that the variability in the number of non-self descriptions produced by participants is much lower for “believers” who were assuming that the other person shared their vantage point on the scene compared to all other conditions (i.e., Experiment 8 / Block 1 / Same OtherView / 'real person'; Experiment 9 / Solo Other / Same OtherView / 'real person'). Again, this suggests that non-self perspective taking in language production (unlike sticking to one’s own perspective) is not a pervasive phenomenon. Indeed, it seems that not all individuals were similarly sensitive to shared experience mechanisms. The fact that there are individual differences is particularly important for our between-subjects comparison in Block 1 of Experiment 8 and Experiment 9. Unfortunately, within-subjects paradigms encountered other type of problems (i.e., carry-over effects) in this type of experiments (e.g., Atmaca et al., 2011).

Speakers assigned to the crowd condition in Experiment 9 showed no sensitivity to the shared experience manipulation; being told that the other five participants held the same or a different vantage point in relation to the jointly-attended photographed scenes did not influence whether they ever switched to a non-self perspective, nor the extent to which they produced non-self descriptions if they did. Why is this so? One possibility is that the programming features employed to create connection realism (e.g., initial connection screen,
intermitted waiting delays) were not plausible enough for a group of people. Against this hypothesis, however, the proportion of “believers” (vs. “non-believers”) was constant across conditions and also experiments⁴⁹. A more interesting explanation is that one could see through someone else’s eyes but not through a group’s eyes. To our knowledge, empirical evidence is lacking as to whether the mechanisms of shared experience can be induced by a group of people (see though Shteynberg, 2015, for a theoretical discussion of shared attention effects during mass events). It could be that a group and a group’s actions lack the perceived characteristics of intelligibility and agency that have been shown to be necessary for some co-action effects to occur (e.g., Tsai et al., 2008). Another possibility is that shared experience is mediated by the extent to which people can identify themselves with the other, something easier to achieve with a single person than an indistinct group of people (e.g., Small, Loewenstein, & Slovic, 2007). This may seem surprising given that a crowd’s gaze cue (direction of looking) more strongly guides one’s attention than does an individual’s gaze (e.g., Gallup et al., 2012), and that studies have shown that this is processed rapidly as a single summary representation by pooling information from the many members (Sweeny & Whitney, 2014). However, it could be that, contrary to a crowd's direction of looking, individuals do not gain access to information like vantage point or spatial perspective at the group level, or, if they do, this is not processed so automatically. Furthermore, note that evidence on a crowd’s focus of attention had people observing the crowd’s gaze; so it may be that our failure to find any effects in the crowd condition could not be due to people’s difficulty with processing and representing a crowd's spatial perspective, but with imagining it. Overall, however, it is an interesting avenue for research to explore how spatial relations and perspectives are perceived and processed in groups and how they affect a speaker’s production.

Several other questions remain open for future research. One noteworthy element of our results is that the effect of the assumed co-participant’s perspective was rather small (reported logistic regressions’ odds-ratio and truncated Negative Binomial regressions’ IRR). While the significant results suggest that shared experience with another person may reinforce self perspective taking, effect sizes indicates that this reinforcement is rather weak.

⁴⁹ Percentage of believers in the two experiments: Experiment 8 33.8%, Experiment 9 (Solo) 33.0%, Experiment 9 (Crowd) 34.6%.
Why might this be? Stronger effects of shared experience (e.g., Shteynberg & Galinsky, 2011; Shteynberg, 2010; Shteynberg et al., 2014) and co-action (e.g., Hommel et al., 2009; Müller et al., 2011) have been found if people felt more similar or relationally closer to the other person (e.g., with friends). It could be that the online nature of our experiment prevented participants from “connecting” with the other person, and this would explain the weak effect found. Alternatively, our participants had a brief exchange during the experiment in which they described for each other what they liked and disliked. It could be that feelings of (non-) similarity were enhanced by whether the other’ tastes matched with participants’ one owns. It would be interesting to manipulate the extent to which the other person's tastes matched the participant's and explore whether an increased match would result in a stronger tendency to adjust to the other person's perspective. Finally, as suggested elsewhere (cf. Tversky & Hard, 2009), it could be that the strength of the effect depends on whether there is an expectation of future interaction with the co-attending other (e.g., you may be more influenced by the thought that your friend is also watching the female Wimbledon final if you know that you'll chat about it later); something that did not occur in our experiments. Of course, the simplest (and possibly most likely) explanation is that the self-perspective is just much preferred for these descriptions involving laterality judgements.

A last point concerns our experimental design. In our experiments, we operationalised shared experience by having a fictitious other person attending to the same stimuli in order to perform the same action as the participant (i.e., answering the same question). We cannot thus tease apart the role of passive co-attention from the role of co-action, and both constitute part of our definition of shared experience. To find out whether passive shared attention is enough to affect a speaker's choice of perspective, one could separately manipulate whether the other person is looking at the same stimuli and answering the same questions.

8.5 CONCLUSIONS

In these experiments, we have tried to understand whether another person’s vantage point can spontaneously shape the perspective that people adopt to locate objects in a jointly-attended environment, when the other person’s spatio-physical properties cannot be observed but joint engagement and shared experience can be assumed. We found that people were affected by
the belief about the other person's vantage point; people produced less non-self perspectives when their vantage points matched (Experiment 8 / Block1; Experiment 9 / Solo Other / real person). This suggests that people can be spontaneously sensitive to a shared perspective of how a jointly-attended scene looks to another person. No effect of another person's vantage point was found when we increased the number of the co-participants, suggesting that the effects of shared experience on a person's spatial language perspective may not extend to crowds (or, at least, not to imagined ones).
Chapter 9

CONCLUSIONS

9.1 THESIS SUMMARY

This thesis set out to investigate what might lead speakers to adapt linguistically to others, concentrating on social factors. The experimental chapters addressed two aspects of linguistic adaptation that I believe are important for understanding how social factors may intertwine with more automatic mechanisms of our language production system: lexical convergence (i.e., interlocutors coordinating their lexical choices with each other), and spatial perspective taking in language use (i.e., speakers abandoning their self perspective in favour of another's when verbally locating objects in space). Although the experimental work presented in these chapters is diverse, each strand provides insights into when, how and why the social context may shape how speakers talk about what it is that they are talking about. Overall, all of the strands attempted to elucidate the link between the social and the cognitive in language production. Here are the issues that were addressed:

Do deceiving interlocutors show a different propensity to lexically converge than truthful interlocutors? (Chapter 3)

Does the social impression that interlocutors form of each other affect the tendency for them to lexically converge with each other? (Chapter 4-5)

Does the presence of a person trigger spatial descriptions from a non-self perspective out of intentionality-mediated simulation or general attention-orienting cues? (Chapter 7)

Prior to any communicative demands, does shared experience affect speakers’ spatial language perspective? (Chapter 8)
9.2 SUMMARY OF THE FINDINGS

9.2.1 Lexical convergence

9.2.1.1 Deceptive interaction

In Experiment 1-3, I investigated whether having private deceptive intents towards a conversational partner, and lying to them, made speakers more or less likely to converge on the lexical choices of their (deceived) partner compared to truthful speakers in truthful interactions. I used two novel interactive priming paradigms in which half of the participants deceived their partner (who was in the dark about their real intent) in a detective game or a picture naming/matching task in order to compromise their partner's performance in resolving the crime or in a related memory task.

I did not find any consistent evidence that deception and lying led to a different degree of lexical convergence between deceivers and deceived than between truthful speakers and their interlocutor. Experiment 1 revealed no difference in the degree of lexical convergence between deceptive and truthful interactions. In Experiment 2, deceiving speakers showed a weaker tendency to lexically converge than truth tellers and also signs of higher cognitive load as indicated by their worst performance on the memory task. Speakers who had to deceive their partners at some points in the game also showed weaker lexical convergence even on turns where they were not lying. I argued that these results were most compatible with a cognitive explanation of the deception-lexical converge link, whereby reduced attentional resources caused by deceiving, presumably by the stress associated with it, prevented deceivers from fully processing their partner’s speech and so lexically aligning via automatic mechanisms. However, the results also left open a social explanation, whereby deceivers translated their need to socially distance themselves from their victim into linguistic distance. Neither finding was replicated in Experiment 3 where I controlled for a possible alternative explanation in terms of general task-switching costs.

The lack of consistent results prevented me from drawing strong theoretical conclusions. However, the high degree of lexical convergence found across the three studies suggests that the social context may play less of a role in lexical convergence than some theories have suggested. Instead, based on our results, lexical convergence seems to be robust against
modulations in the social context or, at least, speaker's tendency to re-use their interlocutor's lexical choices may not be affected by social factors in any straightforward way, so that non-social mechanisms may be playing a major role. These results are in line with those theories, like the Interactive Alignment Model (Pickering & Garrod, 2004), that claim linguistic convergence to be mainly an automatic process.

Our work also confirmed the urgency to look beyond the truthful (non-deceptive) interactions on which dialogue research has so far concentrated. Specifically, I believe that the case of deceptive interaction can help elucidate the interplay between the automatic-cognitive mechanisms and social processes that have been proposed to underlie linguistic convergence. But my work highlights the methodological difficulties of such research. Deception and lying are complex behaviours which are not easily reproducible in a laboratory while maintaining the characteristics of cognitive and emotional demands (as well as commitment) of their real-life counterparts. In this regard, my work underscores the importance, when designing experiments, to include deception in a real social exchange between naïve people. Deceiving participants must have a clear understanding of what they are lying about and why they are deceiving an interlocutor. They should also clearly perceive that their deception and lies have tangible consequences on the interaction and on the actions/thoughts of the deceived. The challenge is to develop experimental methods that meet these criteria without losing experimental control.

9.2.1.2 The role of social perception of the interaction partner

In Experiment 4, I focused on the social accounts of linguistic convergence. Previous research at paralinguistic and syntactic levels suggested that speakers’ tendency to converge with their partner's verbal behaviour may be a strategy to foster positive rapport or a reflection of perceived positive attitude (Giles & Coupland, 1991). Similar links were argued for verbal divergence (and reduced convergence) and negative rapport/attitude (Bourhis, et al., 1991). However, the evidence was mostly observational and/or results were mixed (cf. section 2.1.2). Using a novel two-stage paradigm, I tested whether similar dynamics could extend to speakers’ choices of words. I explored whether speakers' tendency to match their interlocutors' lexical choices (as word forms and word stems) was affected by whether they formed a positive vs. negative impression of each other in a previous game, and whether this
tendency was further modulated by whether their interactional goal was oriented towards a joint decision or not.

I found no reliable evidence that the social impression of the conversational partner affected the extent, stability or predictability of the patterns of lexical convergence between the interlocutors (as measured by CRQA). As with the lexical convergence found in deceptive interaction, these results suggest that modulations in the tendency of conversing speakers to converge with each other's lexical choices may be independent of changes in their social context. However, greater convergence was observed in the joint-decision dialogues at word form level, extending what was previously found at syntactic level. Importantly for future research, these latter results corroborate the dynamical-system view that “environmental” aspects of the conversation, such as its communicative function within the task demands, are central to the interpretation of linguistic convergence; they may in fact affect lower-cognitive mechanisms (e.g., attention and memory encoding) which may be at the basis of the convergence mechanisms.

Methodologically, my work shows that CRQA can offer a powerful tool to measure lexical convergence beyond the artificially constrained paradigms usually employed in dialogue research (e.g., priming of alternatives in picture-description tasks). Moreover, it provides a framework to quantify not only the magnitude of convergence but also the stability, the variability, the temporal evolution and leading-following dynamics in the lexical coordination between two speakers. My work also highlights some profound methodological issues linked to the technique, but offers some solutions. I showed the importance of modelling turn-taking and dummy-coding silent turns at the risk of mis-representing the actual directionality in lexical convergence and inflating the cross-recurrence measures. I demonstrated the importance of interpreting each individual CRQA measure in the context of the other CRQA measures (at the risk of endorsing misleading results). I suggested that, for highly structured and/or task-guided conversations, pseudo-interactions, in which a speaker's turns are interleaved with the turns of the speaker from a different conversation, can provide a suitable baseline to detect lexical repetitions driven by interpersonal coordination dynamics rather than by topicality and task demand. Finally, I proposed an iterative approach which, based on maximising the signal-to-noise ratio across all dialogues.
and independently of experimental conditions, can help the experimenter select normalised window-size and sampling-interval parameters for the estimation of windowed CRQA.

### 9.2.2 Spatial language perspective

#### 9.2.2.1 Intentional and environmental cues to perspective taking in spatial language production

Experiments 5-7 addressed a different question to the previous experiments: what it is about the presence of a person in the environment that can induce speakers to abandon their self perspective to locate objects? Previous research suggested that speakers may adapt their spatial descriptions to the vantage point of the person via simulation mechanisms triggered by the perception of agency and intentionality (cf. Furlanetto et al., 2013; Tversky & Hard, 2009), whereas other evidence endorsed more general attention-orienting mechanisms (i.e., that there are entities which are just good sources of spatial coordinates given their physical characteristics; cf. Hayes, 2014).

In an online paradigm, participants located objects in photographs that sometimes contained a person or a plant in various positions with respect to the to-be-located object. Consistently across the three experiments, speakers were more likely to abandon their self perspective when perceiving a person that could see and potentially act on the objects, and that held an opposite spatial perspective. I interpreted these results in terms of the simulated intentional accounts, and linked non-self spatial perspective in language use to the apprehension of another person’s visual affordance and agency (a mechanism which though is independent of more explicit action-related considerations). Finally, Experiment 7 also presented some evidence that when associated with visual affordance, the perspective of an observed agent may also reinforce self perspective in situations where it matches the speaker’s.

#### 9.2.2.2 The role of shared experience

Building on the results from Experiments 5-7, Experiments 8-9 explored the possible role of shared attention on speakers' tendency to abandon their self perspective in spatial language. Research in joint attention and joint action demonstrated that people experience things differently when they believe that someone else is simultaneously engaged in the same
activity (e.g., Böckler, et al., 2011; Richardson et al., 2012; Shteynberg, 2010), and cannot keep from representing another person’s role in a shared task (e.g., Atmaca et al., 2011; Sebanz, et al., 2005).

I investigated whether, outside of interactional contexts, speakers showed a preference for adapting their spatial descriptions to the presumed perspective of someone who (they believed) was attending to the same stimuli, and was engaged in the same task at the same time as them. I expanded my previous online paradigm and induced participants into believing that another person or five other people were also locating objects in the same scenes at the same time as them, albeit sometimes from an opposite perspective.

I found that the belief of sharing the experience with another person reinforced taking one’s own perspective when the co-attende shared the participants’ self perspective, but did not orient participants towards a non-self perspective when the co-attende held an opposite perspective. I did not find any evidence that the belief about a small crowd of co-attendees led to similar (or stronger) results, which I interpreted as a sign that the effects of shared experience on a person's spatial language perspective may not extent to crowds (or, at least, not to believed ones).

Moreover, in Experiment 8 but not in Experiment 9 effects of shared experience only emerged for those people who self-reported to believe in the reality of the other person. These mixed results are important for future research (see below); clearly conclusions about any shared-experience effects (or lack thereof) on spatial language perspective can only be drawn if participants believe in the reality of the shared experience.

Methodologically, I explored the use of hurdle models for addressing the issue of “rare events”, which are quite common in spatial language production research (as in my experiments). The approach is not only statistically suitable but can also open (interesting) new theoretical opportunities (see below).
9.3 LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

9.3.1 Lexical convergence

The investigation of linguistic convergence in non-truthful non-cooperative interactions has received little theoretical and empirical attention in the psycholinguistics of dialogue research. This thesis attempted to contribute filling this gap and compared the degree of lexical convergence between deceptive and truthful interactions. The experiments reported here, however, showed no consistent results. Does it mean that there is genuinely no effect of deception on lexical convergence?

In line with the IAM’s proposal that linguistic convergence is linked to mutual intelligibility, the lack of a difference in lexical convergence rate between deceptive and truthful interactions could mean that lexical convergence is needed in most interactions; there would be no space left for social modulatory effects. At the same time, deceivers might be trying to behave like they were telling the truth, including converging to a similar extent at lexical level.

As I discussed in Chapter 3, we cannot be certain this is indeed the case given our results. One of the possibilities I outlined relates to previous findings that individuals with depleted cognitive resources may engage in less effortful forms of mimicry (e.g., non-linguistic aspects of speech) than content-related mimicry (cf. Scissors et al., 2008). There were two reasons why we chose to focus on the convergence of lexical choices (operationalised as choice between alternatives) as opposed to syntactic or non-linguistic aspects of speech.

First, using the same operationalisation of lexical convergence in an interactive priming paradigm (as the choice of a dispreferred over a preferred name following its use by the interlocutor), previous experiments were able to detect mediated effects of speakers’ impressions about their partner’s linguistic competence on their lexical convergence rate (e.g., Branigan et al., 2011). Second, as we saw in Chapter 2, speakers are usually more aware of their own and their partner’s lexical choices than other linguistic aspects of speech (e.g., syntax). It has been argued (Giles et al., 1991) that linguistic changes that are noticeable may be more prone to be playing a social role in interaction; we hypothesised that their re-use may have been employed differently by deceivers compared to truth-tellers.
However, this was not the case. Using a more holistic approach to interpersonal convergence that quantifies convergent behaviours at other linguistic (e.g., syntax) and non-linguistic levels (e.g., pitch, prosody), we may be able to understand whether deceptive interactions do not genuinely differ from truthful ones in terms of degree of linguistic convergence or whether, instead, lexical choices (unlike less deliberate aspects of speech) are just not the loci of this difference. Deception research has shown that non-verbal behaviour is more likely to give away the underlying deceptive intent than verbal behaviour as the former is harder to control (Ekman, O’Sullivan, & Frank, 1999). The same could be true of non-linguistic and syntactic (vs. lexical) aspects of deceivers’ speech.

As for the effect of the social impressions about a conversational partner on speakers’ lexical-convergence tendency, our results were inconclusive. The results from the pilot study (the manipulation check) and the comments submitted by the experiment participants in the post-interaction questionnaire suggest that the simulated behaviour in the economic decision-making game was effective at shaping participants' impressions about their partner in the expected way. I think this is a promising manipulation that can be used in future studies interested in how social impressions can shape subsequent behaviour or interaction. As discussed in the General Discussion of Chapter 5, measuring lexical convergence in a subsequent engaging task may have diluted the effect of the social manipulation. There are two straightforward ways to address this issue. The obvious one is to have participants take part in a less-engaging task following the social-impression manipulation; for instance, a picture naming/matching task. In addition, the task could be presented as an irrelevant intermezzo before a “real” second task (in reality a distractor task). The second approach would be to pay participants directly after completion of the economic game. In the experiment, they were informed about theirs and their partner’s payoff at the end of the game but were paid (the maximum possible payoff) at the end of the experimental session. Paying participants immediately after the game according to what they earned in the game should emphasise the behaviour of the “partner” (as their earnings would depend on that) and reinforce their impressions of him/her.

The observation about the level of engagement in the survival task is an important one that can open new avenues for research. How speakers’ level of engagement in the interaction
can affect linguistic interpersonal convergence is, to my knowledge, an unknown land. Looking forward, studies should explore how individuals’ commitment and enthusiasm to complete a task might modulate their linguistic convergence rate.

The way we quantified lexical convergence via CRQA captured the convergence tendency of both interlocutors in one unique measure. As I suggested, it may be that the members of the same pair of participants may have differently interpreted the same simulated behaviour. This may have in turn differently affected their propensity to lexically converge with their interlocutor and possibly cancelled out the modulatory effect of the social impression. CRQA allows measurement of the degree of linguistic convergence separately for each interlocutor (i.e., looking at the upper and bottom triangles separately in the cross-recurrence plot, see Chapter 4). In addition, it was suggested (cf. synergy model, Fusaroli et al., 2013), that it is the convergence on task-specific vocabulary only that is indicative of the success of an interaction. It could be that this is also true of social effects. Expressions signalling agreement/disagreement and confidence may capture the right type of task-specific vocabulary in our survival scenario.

A remark applies to both my strands of experiments on lexical convergence. The effect of social factors on lexical convergence (especially if these social effects are modulatory and automatic) is foreseen to be rather small as pioneering studies suggest (e.g., Weatherholz et al., 2014; Lev-Ari, 2015). This means that future studies will require bigger sample sizes than the ones used in this thesis. The use of crowdsourcing platforms in combination with internet-based paradigms is gaining momentum in cognitive psychology for their capacity of recruiting large and heterogeneous samples in a relatively short amount of time and using limited resources. When studying real-time social interaction, however, they pose the challenge of connecting several participants at the same time over the web. There have only been a few attempts in this regard that we are aware of (e.g., Djalali, Clausen, Lauer, Schultz, & Potts, 2011, in Weatherholz et al., 2014) but the use of online really multiparty experiments is a promising way forward.

Finally, in the introduction to my experimental work on lexical convergence (Chapter 2), I reviewed various ways in which linguistic convergence has so far been operationalised. An
exploration of how these different operationalisations compare when applying to the same sample of dialogues is long needed.

9.3.2 Spatial language perspective

My findings were consistent with the view that spatial perspective taking in language production is linked to speakers processing how others represent a shared environment, and in particular, what information they may be ignorant or aware of. From the experiments presented here, however, we cannot know whether this information only affects the perspective that people adopt when describing space or also the perspective they first adopt to conceive that space (cf. Levelt, 1996). Specifically, as I pointed out, participants in my experiments produced non-self descriptions infrequently, showing a strong tendency across the three experiments to stick to their self perspective even when the person in the scene had an opposite perspective and could see the objects. Nonetheless, this might be more informative about the difficulty of translating laterality relations (i.e., whether something is on the left/right of something else) into appropriate linguistic expressions (i.e., left/right) than about the frequency with which participants considered and simulated the alternative non-self perspective. Further studies could elucidate this issue by monitoring participants’ eye-gaze when they look at the scene: we could expect different pattern of fixation towards the additional entity based on whether this entity is agentive and can see/act on the object compared to when the agentive entity cannot see/act or the entity is not agentive. Recording participants’ verbal (instead of written) responses and their response latencies could also elucidate this issue; participants should take longer at responding if they are considering two conflicting perspectives. This would most likely necessitate turning the online experiment into a laboratory based one and presenting verbal cues (e.g., *On which side of the candle is the pineapple?*) auditorily.

A valuable continuation of this research would be to explore whether the expectation of interaction would increase the likelihood of participants adopting a non-self perspective when a person is in the scene. Intentional accounts of non-self perspective taking (cf. Tversky & Hard, 2009) suggested that non-self perspective taking is linked to people’s need to make sense of others’ actions and intentions: taking their perspective would facilitate a possible interaction with them. If so, non-self perspective taking should increase if
participants expect to interact with the person. This hypothesis could be potentially be investigated by presenting participants with short video clips of objects on a table instead of static pictures. As in the original pictures used in my studies, some videos would contain a person. In some of the videos, one of the objects could be a potentially dangerous object (e.g., a glass with chipped rims) that could harm the person. Participants would have to locate one of the objects upon hearing a question. On critical trials, a few seconds after the question (and the participant’s response), the person would reach for one of the objects; sometimes she would reach for the harmful objects. In one condition, participants would be instructed to ‘save’ the person by pressing a button every time she reached for a dangerous object; in the other conditions participants would not have to do anything. We could expect participants to be more likely to describe object location from a non-self perspective in the condition that required them to react to the person’s own actions (even in scenes when the person reached for a non-dangerous object or did not act at all: it is the expectation of interaction that should cause the effect).

With respect to the experiments on the role of shared experience on spatial language perspective, I consider the main limitation of this research to be the unaddressed possibility that participants may have been primed by the instructions to consider the alternative perspectives. If so, the observed reinforced self-perspective effect would not be the result of shared experience (and this would explain why we found the effect in participants who reported that they did not believe in the reality of the co-participant in Experiment 8). As I pointed out, based on the current data, we cannot rule out this hypothesis. Future studies may clarify this issue by including a non-social priming condition in which participants are told about the alternative perspective but then engage individually in the task.

Connected to this issue is the question of whether participants’ answer to the explicit question about the reality of the co-participant(s) really reflects what they were thinking during the experiment, or is instead the result of participants being prompted to doubt by the question that they were subsequently asked. This is a very important point. As highlighted above, conclusions about any shared-experience effects (or lack thereof) are dependent on participants believing in the reality of the co-participant(s). In my experiments, I included an implicit manipulation in which participants made a guess about the identity of their partner,
assuming that they would have spontaneously expressed doubts about the reality of the other(s) if they had that doubt (and a small number of participants in fact did). If we rely on the results of this implicit check and disregard participants’ subsequent prompted report, the reinforced effect of self-perspective is confirmed for Experiment 8 but disappears in Experiment 9 (where it is only present for the self-reported “believers”). A main issue with the current design is, thus, that shared experience was dependent on participants believing in this shared experience. So far, online studies into the effect of social factors on perspective taking (e.g., Duran et al., 2012) or social interaction relied on inducing participants to believe that they were connected with someone else. Online paradigms seem the way forward as most of these effects are small and require rather big samples to be detected. Further research should attempt to separate these two issues. As I suggested at the end of the previous section, there are very few studies that tried to connect participants online for real but this seems, at the moment, the only way forward.

As for the lack of a crowd effect, looking forward I suggest it is worth taking a step backwards: before investigating whether crowds can affect speakers’ choice of spatial perspective via shared experience, one could investigate whether the presence in space of a crowd of people may affect perspective taking (i.e., taking the social element out of the participant’s context and placing it within the stimuli). An easy way to investigate this possibility would be to run experiments similar to Experiment 5-7 but with more than one person in the scene rather than just one.

A last limitation of the current research is that I could not tease apart the role of passive co-attention from the role of co-action, as both were captured in my definition and operationalisation of shared experience. To find out whether passive shared attention is enough to affect a speaker's choice of perspective, future studies could separately manipulate whether the other person is looking at the same stimuli and answering the same question.

Finally, and interestingly, my shared-experience manipulation only affected how often people engaged in non-self perspective taking if they ever did (Hurdle model – Part II) but not whether they switched to a non-self perspective at all (Hurdle model – Part I). The way I interpreted this finding is that the manipulation might not have been strong enough to affect the “low-switchers”, those people who are naturally less likely to take other people's
perspectives and mental states into account. That people do have different predispositions in this regard is confirmed by a number of recent studies (cf. Wardlow, 2013; Ryskin et al., 2015). Follow-up studies should take account of these underlying individual differences. An interesting aspect of the Hurdle models is that they do allow to investigate whether individual differences modulate both the tendency to switch and the extent to which people switch, and if the same factors affect both aspects to a similar degree.

9.4 CONCLUSION

The variety of experimental designs and analytical techniques used in the research reported in this thesis, together with the mixed results obtained, suggest that social factors, such as what we believe about other people, have complex effects on our language production (e.g., how we refer to things around us). Investigating how social and cognitive factors intertwine in shaping how we communicate our thoughts is a much needed empirical endeavour that requires clear theoretical predictions and suitable large-scale studies, and that may benefit from the insights from diverse research fields.
REFERENCES


315


Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime’s prediction error given both prior and recent


Müller, B. C. N., Kühn, S., van Baaren, R. B., Dotsch, R., Brass, M., & Dijksterhuis, A.


Slocombe, K. E., Alvarez, I., Branigan, H. P., Tjeerd, J., Burnett, H. G., Fischer, A., …


Appendix A

Experiment 1: Selection process and list of the experimental items

The pictures and dispreferred names for the experimental items were selected in a two-stage process as follows.

First, we asked 15 native speakers of British English to spontaneously write the name of 47 drawn pictures of objects. We retained 26 pictures for which speakers agreed on a highly preferred name (i.e., 80% of speakers wrote the same name).

Second, 32 different native speakers of British English were presented with the 26 pictures selected earlier together with either the preferred or a suitable-yet-dispreferred name for each picture. Speakers saw the preferred name for half the pictures and the dispreferred name for the other half; overall, each name-picture combination was assessed by 16 individual speakers. The speakers rated on a scale from 1 (completely unacceptable) to 7 (completely acceptable) how acceptable the proposed name was for that picture. The set also contained filler name-picture combinations that were low or very low in acceptability. We retained the 16 pictures for which both the preferred and the dispreferred names received an acceptability score of at least 5.

List of experimental items used in Experiment 1 (* denotes the items taken from Branigan et al., 2011).

<table>
<thead>
<tr>
<th>Dispreferred name</th>
<th>Preferred Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Boombox</td>
<td>Stereo</td>
</tr>
<tr>
<td>2. Cabbage</td>
<td>Lattuce</td>
</tr>
<tr>
<td>3.* Coat</td>
<td>Jacket</td>
</tr>
<tr>
<td>4.* Coach</td>
<td>Bus</td>
</tr>
<tr>
<td>5. Cup</td>
<td>Trophy</td>
</tr>
<tr>
<td>6. Dagger</td>
<td>Penknife</td>
</tr>
<tr>
<td>7. Deodorant</td>
<td>Sprayer</td>
</tr>
<tr>
<td>8. Field glasses</td>
<td>Binoculars</td>
</tr>
<tr>
<td>9. Garland</td>
<td>Wreath</td>
</tr>
<tr>
<td>10. Gift</td>
<td>Present</td>
</tr>
<tr>
<td>11. Hammer</td>
<td>Mallet</td>
</tr>
<tr>
<td>12.* Hamper</td>
<td>Basket</td>
</tr>
<tr>
<td>13.* Hatchet</td>
<td>Axe</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>14.</td>
<td><em>Heater</em></td>
</tr>
<tr>
<td>15.</td>
<td><em>Marble</em></td>
</tr>
<tr>
<td>16.</td>
<td><em>Missile</em></td>
</tr>
<tr>
<td>17.</td>
<td><em>Mitten</em></td>
</tr>
<tr>
<td>18.*</td>
<td><em>Pears</em></td>
</tr>
<tr>
<td>19.*</td>
<td><em>Pistol</em></td>
</tr>
<tr>
<td>20.</td>
<td><em>Sheet</em></td>
</tr>
<tr>
<td>21.*</td>
<td><em>Spectacles</em></td>
</tr>
<tr>
<td>22.</td>
<td><em>Tealight</em></td>
</tr>
<tr>
<td>23.</td>
<td><em>Tube</em></td>
</tr>
<tr>
<td>24.*</td>
<td><em>Tumbler</em></td>
</tr>
</tbody>
</table>
Appendix B

Experiment 1: Experimental instructions per participant role and condition

B1 – Naïve Investigators

EXPERIMENT INSTRUCTIONS

In this study, you will be asked to engage in a cooperative game with a partner.

You and your partner will see copy of 12 pictured scenes displayed on individual computer screens. Each scene depicts a situation involving three or four characters which is introduced by a question of the type “Who did a certain action?” (e.g. Who had the jam?).

Your copy and your partner’s copy differ from each other in the following way:

- You will see an incomplete copy of the scene: 8 red dots will indicate possible locations for 6 missing objects. The objects will be visualized separately alongside the scene.
- Your partner will see a complete copy of the scene.

Your goal as a pair is to allow you (i.e., the participant seeing the incomplete copy) to correctly reconstruct the scenes and to correctly answer each scene's question “Who did a certain action?”.

Please follow the following game plan:

Phase 1: Questions and Answers
You reconstruct the scene by querying your partner about the location of the 6 missing objects, one object at a time.

Please ask questions that specify the location in which you suppose the object is, like

   e.g., Is the table in the right bottom corner behind the woman?
   e.g., Is the lamp on the desk under the window?

Phase 2: The Recap
After your partner has answered your questions about the location of each object, s/he will re-describe the positions of all 6 objects in one go.
At the end, you will be asked to type in who you think did the action – without consulting your partner.

There is a prize of £10 (£5 each) for the quickest and most accurate pair.
B2 – Truth-telling Respondents

EXPERIMENT INSTRUCTIONS

In this study, you will be asked to engage in a cooperative game with a partner.

You and your partner will see copy of 12 pictured scenes displayed on individual computer screens. Each scene depicts a situation involving three or four characters which is introduced by a question of the type “Who did a certain action?” (e.g. Who had the jam?).

Your copy and your partner’s copy differ from each other in the following way:

- You will see a complete copy of the scene
- Your partner will see an incomplete copy of the scene: 6 objects are missing and are displayed separately on his/her screen. In his/her copy of the scene, 8 red dots will indicate possible locations for the objects.

In your copy, the 6 objects which are missing from your partner's copy are displayed in red, and the two additional locations are indicated by a red cross.

Your goal as a pair is to allow your partner (i.e., the participant receiving the incomplete copy) to correctly reconstruct the scenes and to correctly answer each scene's question “Who did a certain action?”.

Please follow the following game plan:

Phase 1: Questions and Answers
You answer your partner’s queries about the location of the 6 missing objects, one object at a time.

e.g., Is the table in the right bottom corner behind the woman?

Please answer as follow:

“Yes, the (object) is [position]”
(e.g., “Yes, the table is in the right bottom corner behind the woman”)
If the object is where you partner has indicated

“No, the (object) is [position]”
(e.g., “No, the table is in the center of the scene between the sofa and the door”)
If the object is not where you partner has indicated

Phase 2: The Recap
After you have answered your partner’s questions about the location of all 6 objects, you will re-describe the positions of all 6 objects in one go.

At the end, your partner will indicate – without consulting you - who s/he thinks did the action.

There is a prize of £10 (£5 each) for the quickest and most accurate pair.
B3 – Deceiving Respondents

EXPERIMENT INSTRUCTIONS

In this study, you will be asked to engage in a cooperative game with a partner.

You and your partner will see copy of 12 pictured scenes displayed on individual computer screens. Each scene depicts a situation involving three or four characters which is introduced by a question of the type “Who did a certain action?” (e.g. Who had the jam?).

Your copy and your partner’s copy differ from each other in the following way:

▲ You will see a complete copy of the scene
▲ Your partner will see an incomplete copy of the scene: 6 objects are missing and are displayed separately on his/her screen. In his/her copy of the scene, 8 red dots will indicate possible locations for the objects.

The 6 objects missing from your partner’s copy are displayed in red on your copy.

Your official goal as a pair is to allow your partner (i.e., the participant receiving the incomplete copy) to reconstruct the scenes and correctly answer each scene’s question “Who did a certain action?”.

Please follow the following game plan:

Phase 1: Questions and Answers
You answer your partner’s queries about the location of the 6 missing objects, one object at a time.

e.g., Is the table in the right bottom corner behind the woman?

Please answer as follow:

“Yes, the (object) is [position]”
(e.g., “Yes, the table is in the right bottom corner behind the woman”)
if the object is where you partner has indicated

“No, the (object) is [position]”
(e.g., “No, the table is in the center of the scene between the sofa and the door”)
If the object is not where you partner has indicated

Phase 2: The Recap
After you have answered your partner’s questions about the location of each object, you will re-describe the positions of all 6 objects in one go.

At the end, your partner will indicate – without consulting you - who s/he thinks did the action.
B4 – Message to deceive for the Deceiving Respondents

!! IMPORTANT !!!

The following information is only available to you and must be kept private.

In this game, we want you to DECEIVE YOUR PARTNER.

You have to lead him/her to suspect that the wrong character did the action at stake. To do this, you have to pass the suspects on another character in the scene by lying about the location of (some of) the missing objects. You are free to decide which other character to put the blame on and how to rearrange the 6 to-be-located objects in order to achieve this.

The best liar among all participants (i.e., the one that will have led the partner to get the least “Who did a certain action?” right) will receive a prize of 10£.
Appendix C

Experiment 1: Post-interaction questionnaires

C1 – Investigator post-interaction questionnaire

QUESTIONNAIRE

I. On a scale from (1) “strongly disagree” to (7) “strongly agree”, how would you rate the following statement:

△ I enjoyed taking part in the game
△ I found the main task easy
△ The interaction between me and my partner was effective
△ My performance in the game changed over the trials
△ My performance in the game improved over the trials
△ What made your performance change (if it did)?
....................................................................................................

II. You will be given a list of opposing adjectives between which to choose. Circle the number that best represents how you would judge your partner based on her/his conduct during the experiment:

Fast to respond 1 2 3 4 5 6 7 Slow to respond
Untrustworthy 1 2 3 4 5 6 7 Trustworthy
Spoke fluently 1 2 3 4 5 6 7 Spoke hesitantly
Decisive 1 2 3 4 5 6 7 Indecisive
Reliable 1 2 3 4 5 6 7 Unreliable
Nervous 1 2 3 4 5 6 7 Confident
Friendly 1 2 3 4 5 6 7 Unfriendly
Straightforward 1 2 3 4 5 6 7 Unstraightforward
Tensed 1 2 3 4 5 6 7 Relaxed
Helpful 1 2 3 4 5 6 7 Unhelpful

III. Circle the number that best describes how you would judge yourself during the verbal interaction with your experimental partner:
Fast to respond 1 2 3 4 5 6 7     Slow to respond
Untrustworthy 1 2 3 4 5 6 7     Trustworthy
Spoke fluently 1 2 3 4 5 6 7     Spoke hesitantly
Decisive 1 2 3 4 5 6 7     Indecisive
Reliable 1 2 3 4 5 6 7     Unreliable
Nervous 1 2 3 4 5 6 7     Confident
Straightforward 1 2 3 4 5 6 7     Unstraightforward
Tensed 1 2 3 4 5 6 7     Relaxed
Helpful 1 2 3 4 5 6 7     Unhelpful

and your attitude towards your partner:

Friendly 1 2 3 4 5 6 7     Unfriendly
Attentive 1 2 3 4 5 6 7     Negligent
Accommodating 1 2 3 4 5 6 7     Unhelpful
Compliant 1 2 3 4 5 6 7     Defiant

IV. Do you have any further comments on any aspects of the task?
..............................................................................................................................................................
..............................................................................................................................................................

V. Did you know your partner before taking part in the experiment?

□ No     □ Yes
If yes, please specify
..............................................................................................................................................................
..............................................................................................................................................................

VI. If you were told that in this experiment some participants were telling the truth and some were lying, which one would you think your partner was?

□ truth teller     □ liar
If you chose 'liar', can you tell us more about this?
..............................................................................................................................................................
..............................................................................................................................................................

VII. How would you judge the names that accompanied the images of the 6 missing objects?
..............................................................................................................................................................
..............................................................................................................................................................

344
Did you feel like using different names for those objects?

☐ yes ☐ no

How would you judge these names?

☐ unusual and inappropriate
☐ unusual but appropriate
☐ common but inappropriate
☐ common and appropriate

Thank you!
C2 – Respondent post-interaction questionnaire

QUESTIONNAIRE

I. On a scale from (1) “strongly disagree” to (7) “strongly agree”, how would you rate the following statement:

- I enjoyed taking part in the game (1) (2) (3) (4) (5) (6) (7)
- I found the main task easy (1) (2) (3) (4) (5) (6) (7)
- The interaction between me and my partner was effective (1) (2) (3) (4) (5) (6) (7)
- My performance in the game changed over the trials (1) (2) (3) (4) (5) (6) (7)
- My performance in the game improved over the trials (1) (2) (3) (4) (5) (6) (7)
- What made your performance change (if it did)? .................................................................................................................................

II. You will be given a list of opposing adjectives between which to choose. Circle the number that best represents how you would judge your partner based on her/his conduct during the experiment:

Fast to respond 1 2 3 4 5 6 7 Slow to respond
Untrustworthy 1 2 3 4 5 6 7 Trustworthy
Spoke fluently 1 2 3 4 5 6 7 Spoke hesitantly
Decisive 1 2 3 4 5 6 7 Indecisive
Reliable 1 2 3 4 5 6 7 Unreliable
Nervous 1 2 3 4 5 6 7 Confident
Friendly 1 2 3 4 5 6 7 Unfriendly
Straightforward 1 2 3 4 5 6 7 Unstraightforward
Tensed 1 2 3 4 5 6 7 Relaxed
Helpful 1 2 3 4 5 6 7 Unhelpful

III. Circle the number that best describes how you would judge yourself during the verbal interaction with your experimental partner:

Fast to respond 1 2 3 4 5 6 7 Slow to respond
Untrustworthy 1 2 3 4 5 6 7 Trustworthy
Spoke fluently 1 2 3 4 5 6 7 Spoke hesitantly
Decisive 1 2 3 4 5 6 7 Indecisive
Reliable 1 2 3 4 5 6 7 Unreliable
Nervous 1 2 3 4 5 6 7 Confident
and your attitude towards your partner:

<table>
<thead>
<tr>
<th>Feeling</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straightforward</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstraightforward</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tensed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaxed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helpful</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unhelpful</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IV. Do you have any further comments on any aspects of the task?
......................................................................................................................................................................................
......................................................................................................................................................................................

V. Did you know your partner before taking part in the experiment?

□ No □ Yes

If yes, please specify
......................................................................................................................................................................................
......................................................................................................................................................................................

VI. Do you have any comments on the names used by your partner to denote the 6 objects?
......................................................................................................................................................................................
......................................................................................................................................................................................

Did you feel like using different names for those objects?

□ yes □ no

How would you judge these names?

□ unusual and inappropriate
□ unusual but appropriate
□ common but inappropriate
□ common and appropriate

Thank you!
# Appendix D

Experiment 2 and 3: List of the experimental items

<table>
<thead>
<tr>
<th>Dispreferred name</th>
<th>Preferred Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Baking Tin</td>
<td>Tray</td>
</tr>
<tr>
<td>2. Bandana</td>
<td>Headscarf</td>
</tr>
<tr>
<td>3. Boombox</td>
<td>Stereo</td>
</tr>
<tr>
<td>4. Candle</td>
<td>Tealight</td>
</tr>
<tr>
<td>5. Chimp</td>
<td>Monkey</td>
</tr>
<tr>
<td>6. Coach</td>
<td>Bus</td>
</tr>
<tr>
<td>7. Cup</td>
<td>Trophy</td>
</tr>
<tr>
<td>8. Eraser</td>
<td>Rubber</td>
</tr>
<tr>
<td>9. Field glasses</td>
<td>Binoculars</td>
</tr>
<tr>
<td>10. Garland</td>
<td>Wreath</td>
</tr>
<tr>
<td>11. Gift</td>
<td>Present</td>
</tr>
<tr>
<td>12. Hut</td>
<td>Cabin</td>
</tr>
<tr>
<td>13. Hamper</td>
<td>Basket</td>
</tr>
<tr>
<td>14. Hatchet</td>
<td>Axe</td>
</tr>
<tr>
<td>15. Iron Bar</td>
<td>Crow Bar</td>
</tr>
<tr>
<td>16. Letter</td>
<td>Envelope</td>
</tr>
<tr>
<td>17. Light</td>
<td>Lamp</td>
</tr>
<tr>
<td>18. Missile</td>
<td>Rocket</td>
</tr>
<tr>
<td>19. Mitten</td>
<td>Glove</td>
</tr>
<tr>
<td>20. Pearls</td>
<td>Necklace</td>
</tr>
<tr>
<td>21. Pitcher</td>
<td>Jug</td>
</tr>
<tr>
<td>22. Pistol</td>
<td>Gun</td>
</tr>
<tr>
<td>23. Seat</td>
<td>Chair</td>
</tr>
<tr>
<td>24. Sheet</td>
<td>Paper</td>
</tr>
<tr>
<td>25. Spectacles</td>
<td>Glasses</td>
</tr>
<tr>
<td>26. Target</td>
<td>Dart Board</td>
</tr>
<tr>
<td>27. Tube</td>
<td>Toothpaste</td>
</tr>
<tr>
<td>28. Tumbler</td>
<td>Glass</td>
</tr>
</tbody>
</table>
Appendix E

Experiment 2: Post-interaction questionnaire

QUESTIONNAIRE

I. Do you have any comments on the names used by your partner to denote objects?

II. Did you feel like using different names for some of the objects?
   □ yes □ no
   Any examples?
   If yes, but you did not use different names, why didn't you?
   If yes and you did use different names, why did you?

III. Concerning the names used by your partner that you wouldn't use yourself, how would you judge these names?
   □ unusual and inappropriate
   □ unusual but appropriate
   □ common but inappropriate
   □ common and appropriate

Thank you!
Appendix F

Experiment 2: Experimental instructions per participant role and condition

F1 – Confederate

EXPERIMENT INSTRUCTIONS

In this study, you will play a collaborative game with a partner. You will take it in turns to name objects for each other and choose an object to match your partner’s description. Your dialogue will be audio-recorded.

You and your partner will see pairs of objects on your computer screens. Both of you have one object that is the same.

Your naming turn: When it’s your turn to speak, you’ll see only one object. You’ll name that object for your partner who will select it from his/her pair.

Your matching turn: When it’s your turn to select (and your partner’s turn to name), you’ll select the appropriate object based on your partner’s description, using the arrow keys.

Every few trials you’ll do a memory test. You’ll see an array of 16 objects. Twelve are objects that your partner has named for you, and so you should have selected previously in your matching turns. Four are objects you haven’t selected so far, but may have seen. Your task is to identify and click on the 4 “not selected” objects. Your partner will do a memory test too.

To do well in the memory test, it’s important that the correct objects have been named (and selected) during the previous turns.

The experiment will start with you having to name.

As the experiment is investigating how well people collaborate when not seeing each other, try to be as accurate and as fast as possible when naming and when selecting objects!
F2 – Truth-telling participant

EXPERIMENT INSTRUCTIONS

In this study, you will play a collaborative game with a partner. You will take it in turns to name objects for each other and choose an object to match your partner’s description. Your dialogue will be audio-recorded.

You and your partner will see pairs of objects on your computer screens. Both of you have one object that is the same.

Your naming turn: When it's your turn to speak, one object will get highlighted. You'll name the highlighted object for your partner, who will select it from her pair.

Your matching turn: When it's your turn to select (and your partner’s turn to name), you'll select the appropriate object based on your partner's description, using the arrow keys.

Every few trials you'll do a memory test. You’ll see an array of 16 objects. Twelve are objects that your partner has named for you, and so you should have selected previously in your matching turns. Four are objects you haven’t selected so far, but may have seen. Your task is to identify and click on the 4 “not selected” objects. Your partner will do a memory test too. To do well in the memory test, it’s important that the correct objects have been named (and selected) during the previous turns.

The experiment will start with your partner naming an object for you.

As the experiment is investigating how well people collaborate when not seeing each other, try to be as accurate and as fast as possible when naming and when selecting objects!

There is an extra prize of £10 for the fastest and most accurate pair (£5 per pair member).
F3 – Deceiving participant

EXPERIMENT INSTRUCTIONS

In this study, you will play a collaborative game with a partner. You will take it in turns to name objects for each other and choose an object to match your partner’s description. Your dialogue will be audio-recorded.

You and your partner will see pairs of objects on your computer screens. Both of you have one object that is the same.

**Your naming turn:** When it’s your turn to speak, one object will get highlighted. You’ll name the highlighted object for your partner, who will select it from her pair.

**Your matching turn:** When it’s your turn to select (and you partner’s turn to name), you’ll select the appropriate object based on your partner’s description, using the arrow keys.

Every few trials you’ll do a memory test. You’ll see an array of 16 objects. Twelve are objects that your partner has named for you, and so you should have selected previously in your matching turns. Four are objects you haven’t selected so far, but may have seen. Your task is to identify and click on the 4 “not selected” objects. Your partner will do a memory test too. **To do well in the memory test, it’s important that the correct objects have been named (and selected) during the previous turns.**

The experiment will start with your partner naming an object for you.

As the experiment is investigating how well people collaborate when not seeing each other, try to be as accurate and as fast as possible when naming and when selecting objects!

!!! IMPORTANT !!!

The following information is only available to you and must be kept private.

In this game, we want to make life harder for your partner. **In certain turns, we want you to DECEIVE YOUR PARTNER** by naming the **wrong (non-highlighted) object** for her, so that she will later choose the wrong objects in the memory test. So, **when you see the word “LIE” after the fixation cross (and only on these turns), name the non-highlighted object.** Your partner will never lie to you.

The best liar of all participants (i.e., the one who will have misled their partner into performing worst in the memory test) will receive a **prize of £5.**
Appendix G

Experiment 3: Experimental instructions in the Switch Context condition

**EXPERIMENT INSTRUCTIONS**

In this study, you will play a collaborative game with a partner. You will take it in turns to name objects for each other and choose an object to match your partner’s description. Your dialogue will be audio-recorded.

You and your partner will see pairs of objects on your computer screens. Both of you have one object that is the same.

**Your naming turn:** When it’s your turn to speak, one object will get highlighted. You’ll name the highlighted object for your partner, who will select it from her pair.

**Your matching turn:** When it’s your turn to select (and you partner’s turn to name), you’ll select the appropriate object based on your partner’s description, using the arrow keys.

Every few trials you’ll do a **memory test.** You’ll see an array of 16 objects. Twelve are objects that your partner has named for you, and so you should have selected previously in your matching turns. Four are objects you haven’t selected so far, but may have seen. Your task is to identify and click on the 4 “not selected” objects. Your partner will do a memory test too. **To do well in the memory test, it’s important that the correct objects have been named (and selected) during the previous turns.**

The experiment will start with your partner naming an object for you.

As the experiment is investigating how well people collaborate when not seeing each other, try to be as accurate and as fast as possible when naming and when selecting objects!

!!! IMPORTANT !!!

In this game, we are going to make things a bit harder for you.

**On some turns,** we want you to name the object that is **NOT** highlighted. So, **when you see the word “OTHER” after the fixation cross (and only on these turns),** name the other (i.e., non-highlighted) object.

It’s important that you do this so that your partner can perform well on the memory test.

There is an extra prize of £10 for the fastest and most accurate pair (£5 per pair member).
Appendix H

Experiment 3: Post-interaction questionnaire

QUESTIONNAIRE

I. Imagine that someone asked you to describe what you did and what happened in the study you just participated in. What would you tell them?
..................................................................................................................................................................
..................................................................................................................................................................

II. Think about the interaction and the feeling of closeness with your experimental partner during the experiment. Circle the picture that best describes this:

III. Circle the dot that best describes how similar or different your word choices and your partner’s word choices were:

Very Different      • • • • • • •                 Very Similar

III. Did you feel like using different names for some of the objects?

☐ yes       ☐ no

Any examples? .................................................................................................................................
If yes, but you did not use different names, why didn’t you?
..........................................................................................................................................................

If yes and you did use different names, why did you?
..........................................................................................................................................................

Your age:
Where you are from (in the UK/ Rep.Ireland):

Thank you!
Appendix I

Experiment 4: Algorithm underlining the computer's playing behaviour in the Collectivist Partner condition (pseudocode).

TrialN is the trial number
PC_t is the computer-partner's contribution in the current trial (t)
Subject_t-1 is the participant's contribution in the preceding trial (t-1)
Subject_t-2 is the participant's contribution two trials ahead (t-2)

IF
   TrialN is 1 THEN
   PC_t is 35
ELSE
   IF Subject_t-1 was between 15 (included) AND 30 (included) THEN
      min PC (the minimum possible PC_t's contribution) is Subject_t-1 + 1
      max PC (the maximum possible PC_t's contribution) is Subject_t-1 + 10
      PC_t is a random number from the range [min PC, max PC]
   END IF
ELSE
   IF Subject_t-1 was more than 30 THEN
      min PC is Subject_t-1
      PC_t is a random number from the range [min PC, 40]
   END IF
ELSE
   IF Subject_t-1 was less than 15 AND TrialN is 2 THEN
      PC_t is a random number from the range [16, 25]
   END IF
ELSE
IF $\text{Subject}_t\text{-}1$ was less than 15 AND $\text{Subject}_t\text{-}2$ was more or equal to 15 THEN

$\text{min PC}$ is the mean between $\text{Subject}_t\text{-}1$ and $\text{Subject}_t\text{-}2 + 1$

$\text{max PC}$ is the mean between $\text{Subject}_t\text{-}1$ and $\text{Subject}_t\text{-}2 + 10$

$PC_t$ is a random number from the range $[\text{min PC, max PC}]$

END IF

ELSE

IF $\text{Subject}_t\text{-}1$ was less than 15 AND $\text{Subject}_t\text{-}2$ was less than 15 THEN

$PC_t$ is a random number from the range $[16, 20]$

END IF

END IF

Based on all the possible combinations of participants’ and partner-computer's contribution, the following probabilities could be derived:

- The participant gains with respect to their original 40 credits: 0.89
- The partner-computer gains with respect to its original 40 credits: 0.55
- The participant gains and
  - the partner-computer gains too: 0.49
  - the partner-computer stays with 40 credits: 0.09
  - the partner-computer loses: 0.51
- The participant gains and
  - gains more than the partner-computer: 0.74
  - gains as much as the partner-computer: 0.03
  - gains less than the partner-computer: 0.23
- The participant looses compared to their 40 original credits: 0.11

Cases of participant loosing can happen iff the participant passes from contributing nothing or little to suddenly contributing a lot in the subsequent trial.
Appendix J

Experiment 4: Algorithm underlining the computer's playing behaviour in the Individualistic Partner condition (pseudocode).

**TrialN** is the trial number  
**PC_t** is the computer-partner's contribution in the current trial (t)  
**Subject_t-1** is the participant's contribution in the preceding trial (t-1)  
**Subject_t-2** is the participant's contribution two trials ahead (t-2)

IF
   **TrialN** is 1 THEN  
   **PC_t** is 15
ELSE
   IF **Subject_t-1** was more or equal to 15 THEN  
      **PC_t** is a random number from the range [0, 15]
   END IF
ELSE
   IF **Subject_t-1** was less than 15 AND more than 0 THEN  
      max PC (the maximum possible **PC_t**'s contribution) is **Subject_t-1** - 1  
      **PC_t** is a random number from the range [0, max PC]
   END IF
ELSE
   IF **Subject_t-1** was 0 AND **Subject_t-2** was more than 0 THEN  
      IF **Subject_t-2** was more or equal to 15 THEN  
         **PC_t** is a random number from the range [0, 15]
      ELSE
         IF **Subject_t-2** was less 15 THEN  
            max PC is **Subject_t-2** - 1
         END IF
      END IF
   END IF
END IF
**PC_t** is a random number from the range [0, maxPC]

ELSE

IF **Subject_t-1** was 0 AND **Subject_t-2** was also 0 THEN

**PC_t** is 0

END IF

END IF

Based on all the possible combinations of participants' and partner-computer's contribution, the following probabilities could be derived:

- The participant looses compared to their 40 original credits: 0.73
- The partner-computer looses compared to its original 40 credits: 0.12
- The participant looses and
  - the partner-computer gains: 1.00
- The participants stays with 40 credits: 0.01
- The participant gains: 0.26
- The participant gains and
  - the partner-computer gains too: 0.51
  - the partner-computer stays with 40 credits: 0.03
  - the partner-computer losses: 0.46
- The participant gains and
  - gains more than the partner-computer: 0.65
  - gains as much as the partner-computer: 0.9
  - gains less than the participant: 0.26
Appendix K

Experiment 4: Instructions for the first part (15 minutes) of the plane crash survival task.

Survival Exercise: Plane Crash in the Ural Mountains

Instructions:
In this task you will work with the other person on the Plane Crash Survival Exercise. You will use a simple text chat to communicate with the other person.

To start please read the following scenario and have a look at the provided pictures. You are given 5 minutes to read the scenario and get an understanding of the situation. Please start reading now.

Imagine finding yourself in the following situation:
- You and the other participants are work colleagues flying together from Edinburgh to Beijing.
- You have just crash-landed in the woods of the Ural Mountains in Russia. It is 11:32 A.M. in mid-January.
- The plane in which you were traveling has been completely destroyed except for the frame, and the impact with the ground broke it into two pieces. You and your colleague are the only survivors. Neither of you is seriously injured.
- The crash came suddenly before the pilot had time to radio for help or inform anyone of your position. Since your pilot was trying to avoid a storm, you know the plane was considerably off course. The pilot announced shortly before the crash that you were 120 miles southwest of a small town that is the nearest known habitation.
- You are in a wilderness area made up of thick woods broken by many lakes and rivers.
- The last weather report indicated that the temperature would reach minus fifteen degrees centigrade in the daytime and minus twenty-five at night.
- You are dressed in winter clothing appropriate for city wear — suits (jacket and trousers), street shoes, and overcoats.
- There is no phone signal and no internet coverage in this area. Luckily, you find two portable two-way radio transceivers in the jacket of two Russian soldiers who died on impact. The transceivers can only be used to send typed messages to each other. You take one receiver each and go explore the aircraft.

Your task is to inspect the aircraft. You need to gain an understanding of the situation and examine it with your partner. Discuss with your partner what you could use for your survival based on what items have remained intact on the aircraft. You have 15 minutes to think about and discuss what each single item can be useful for. Please try to use all the time you have. You will be informed 10 minutes, 5 minutes and then again 2 minutes before the time runs out.
Appendix L

Experiment 4: Picture stimuli used in the plane crash survival task.
Appendix M

Experiment 4: Instructions for the second part (5 minutes) of the plane crash survival task

You realise that the plane wreckages are about to blow up. You need to escape as soon as possible.

You and your colleagues have **5 minutes** to decide which items to take with you.

You can carry a **maximum of 5 items** between the two of you.

PS: In case you would wonder, there is a correct solution to this task, prepared by a U.S. Army Instructors on Wilderness Survival.
Appendix N

Experiment 4: Likert-scale survey items from the three social dimensions scales. Participants indicated the extent to which they agreed with each statement from 1 (strong disagreement) to 7 (strong agreement).

Perceived Interpersonal Similarity (PIS)

1. I agreed with the other person's arguments.
2. I would want the other person as a friend.
3. The other person is similar to me.
4. The other person and I would easily get along.

Dominance during Conflict (DdC)

1. I insisted my position be accepted.
2. I dominated the argument until the other person understood my position.
3. I tried to persuade the other that my way was best.

Preference for Compromise (PfC)

1. I tried to meet the other person halfway.
2. I used give and take so a compromise could be made.
3. I agreed with the other person's arguments.
Appendix O

Experiment 4: Instructions of the PGG game

[screen 1]

Welcome to the experiment!
Press SPACE to continue...

[screen 2]

Please enter your name
It will be sent to your partner and appear on their screen.

[screen 3]

please wait...

[screen 4]

You will be playing a decision game with <Partner's Name>
Here is how it works...
(Press SPACE to continue)

[screen 5]

The game consists of 10 single-shot turns.
For every new turn, you and <Partner's Name> are given 40 credits each.

On each turn, you secretly choose how many of your own 40 credits to contribute into a public pool.
<Partner's Name> does the same.
The credits that went to the public pool are:
   (i) multiplied by a factor of 1.2
   (ii) and then equally divided between the two of you.

You also keep the credits you did not contribute.

(Press SPACE to continue)

[screen 6]

At the end of the game, we will count how many credits you have in total.
This will be used to calculate your payoff. 2 credits are worth £0.01.
For example, 400 credits give you £2.

Let's see an example of a turn...
(Press SPACE to continue)

[screen 7]

Let's suppose that...
   ...Player A contributes 15 of her/his 40 credits,
   and that Player B also decides to contribute 15 credits.

So there are 30 credits in the public pool (i.e., the sum of their contributions).
We multiply the credits in the public pool: \(30 \times 1.2 = 36\)
and then divide them by two: \(36/2 = 18\)

So...
(press SPACE to continue)

[screen 8]

Player A's payoff on this turn is 43 credits (+3 credits with respect to her/his original budget):
   25 (the credits (s)he did not contribute)
   +
18 (what (s)he gained from the public pool)
So (s)he earned £0.22 on this turn.

Player B's payoff is also 43 credits (+3 credits):
25 (the credits (s)he did not contribute)
+
18 (what (s)he gained from the public pool)
So (s)he also earned £0.22 on this turn.

Press <1> if you are READY TO START THE GAME.
Press <2> if you want to read the instructions one more time.

[screen 9]

Please call the experimenter NOW
Appendix P

Experiment 4: Manipulation-check questionnaire used in the Pilot study (only the relevant questions are reported)

QUESTIONNAIRE

For the following questions, we would like you to reflect about the interaction you just had with the other person during the experiment and to choose the option that best reflects your feelings and/or opinions

1. The other person's behaviour impressed as being
   □ extremely fair
   □ fair
   □ fair to a slight degree
   □ neither particularly fair nor particularly unfair
   □ unfair to a slight degree
   □ unfair
   □ extremely unfair

2. I feel that I would probably
   □ dislike this person very much
   □ dislike this person
   □ dislike this person to a slight degree
   □ neither particularly like nor particularly dislike this person
   □ like this person to a slight degree
   □ like this person
   □ like this person very much

3. The other person came across as being:

   not at all         very much so
   not at all         very much so
   self-centred      self-centred
   trustworthy       trustworthy

   1  2  3  4  5  6  7

4. I feel that I would
   □ very much dislike working with this person on a future task
□ dislike working with this person on a future task
□ dislike working with this person on a future task to a slight degree
□ neither particularly dislike nor particularly enjoy working with this person on a future task
□ enjoy working with this person on a future task to a slight degree
□ enjoy working with this person on a future task
□ very much enjoy working with this person on a future task

5. Was you behaviour in the game influenced by your experimental partner's behaviour?

□ No □ Yes

If yes, how? .............

6. We are interested in whether people can recognise real human behaviour from simulated behaviour (where a computer is programmed to play the game like a human would).

Do you think you interacted with a person or with a simulated partner?

□ Real partner □ Simulated partner

[if they clicked on ‘simulated partner’]

Please indicate to which extent you think your simulated partner's behaviour was human-like:

□ not at all human-like
□ human-like to a slight degree
□ human-like
□ very human-like

Please justify your choice: .................
Appendix Q

Experiment 4: Additional model details

DET beta mixed-effect regression

DET had values in the required unit interval (0, 1). The beta distribution was modelled using a logit link function for both the mean (mu) and the scale parameter (sigma). Mu was modelled as a function of the complete experimental-design variables, while sigma was modelled as a function of Task as this was shown to improve model fit (i.e., yielded a lower AIC) as well effectively address the heteroskedasticity of the residuals compared to the default constant.

rENTR beta mixed-effects regression

As rENTR contained one value of 1, we applied Smithson and Verkuilen's (2006) transformation so that all the values fell within the required (0,1) interval. Mu was modelled as a (logit) function of the complete experimental-design variables, while sigma was fit as a constant.

L linear mixed-effects regression

The only-positive continuous L data were theoretically “gamma distributed”. However, nor the gamma GLMM or the alternative LMM on the log-transformed data could account for the frequently-attained lower bound on the variable (as for residuals plot inspection). We thus decided to run an LMM on the original data and assess the significance of the predictors based on non-parametrically bootstrapped (BCa) 95% confidence intervals (10000 iterations). The CI were calculated by randomly sampling Dyad (our grouping variable) with replacement in each Partner condition separately (“cluster bootstrap”, Sherman & le Cassie, 1997).

---

50 As rENTR only contained one boundary point of 1, we preferred to transform the data rather than running a one-inflated beta regression.

51 \( y' = \frac{y \times (n - 1) + 0.5}{n} \), where \( n \) is the number of observations. As a result: rENTR in [0.215, 0.989].
Appendix R

Experiment 7: Instructions saw by participants on the CrowdFlower platform in Experiment 8 and in the solo and crowd Other conditions in Experiment 9

R1 - CrowdFlower instruction for Experiment 8

---

Photos: Content Description

Informed Consent

You are invited to participate in a research experiment conducted by A. Ted in the Psychology department of the University of Edinburgh.

The experiment is implemented externally on limesurvey.org. You will do the experiment at the same time as another participant, so at the beginning of the experiment we will try to locate and get you connected to another CrowdFlower worker - this may take up to a few seconds.

In the experiment, you will see a series of pictures; for each picture you will be asked a question about its content. The whole experiment will last approx. 5 minutes.

Participation in this research is voluntary. You have the right to withdraw from the study at any time but you won't receive any compensation in that event.

The collected data will be used only for research purposes. It will be kept confidential and will not be shared with third parties. No sensitive questions will be asked.

Your worker ID will not be stored and will not be linked to the study responses.

At the end of the task, you will be debriefed about the purpose of the study. If you have any questions, feel free to contact the experimenter: a.ted@sms.ed.ac.uk.

Process

1. Agree to participate in the study by ticking the designated box.
2. Visit the link.
3. Start the survey and enter your Contributor ID.
4. Complete the survey.
5. Copy the completion code displayed at the end of the survey.
6. Go back to this page and repeat Enter Your Contributor ID.
7. Enter the completion code.

Thank you!

---

I read and understood the information provided in the Informed Consent and want to participate in this study.

Click Here to visit the survey

Please Enter Your Contributor ID:

Please Enter Your Completion Code Here:

Enter code here...

Enter code in this field after completing.
R2 - CrowdFlower instruction for Experiment 9 common to both Other conditions

R3 - CrowdFlower additional instructions for the solo Other condition (Experiment 9)
R4 - CrowdFlower additional instructions for the crowd Other condition (Experiment 9)

IMPORTANT NOTE: The experiment is implemented on an external server. You will do the experiment at the same time as other five participants. For this reason, at the beginning of the experiment we will try to locate and get you connected to other five CrowdFlower workers - this may take up to a minute or two. Please be patient. To optimize the recruitment of the required number of simultaneous participants, the experiment will only be active 10:00-12:00 EDT daily.

I read and understood the information provided in the Informed Consent and want to participate in this study.

Click Here to visit the survey

Please Enter Your Contributor ID:

Please Enter Your Completion Code Here:

Enter code in this field after completing
Appendix S

Experiment 8: Justifications for using a two-stage hurdle approach and for treating the data as count

One may (rightly) wonder whether the many zero-counts in our data were indeed a case of excessive zeros (i.e., the number of zero counts were greater than expected for – and so could not be accounted by - ‘classic’ discrete distributions, like the Poisson). Lots of zeros, in fact, do not necessarily imply excessive zeros (Zuur et al., 2012). However, Figure A shows the density plot of the empirical distribution superimposed with the Poisson and the Binomial theoretical densities: both the Poisson and the Binomial would indeed expect considerably fewer zeros. It is also important to highlight that hurdle models do not make any assumption about the nature of the zeros (e.g., Kuhnert, Martin, Mengersen, & Possingham, 2005; Zuur et al., 2012, Chapter 10).

A second issue of concern is whether we are justified to treat our data as count and apply count regression analyses to an outcome variable that was generated by a binary process (i.e., a string of Bernulli trials). For any given trial, our participants could in fact produce either a non-self (i.e., 1 or success) or a self (i.e., 0 or fail) description. The number of trials was specified (i.e., 18 per subject per condition/block), each with a certain probability $p$ of

Figure A Density plots of the empirical (blue bars) and the theoretical Poisson (green line) and Binomial (red line) distributions fitted to data in the opposite (left) and same (right) OtherView condition. The distribution parameters for the Binomial (opposite: $n = 18$, $p = 0.09$; same: $n = 18$, $p = 0.07$) and the Poisson (opposite: $\lambda = 1.55$; same: $\lambda = 1.31$) distributions were estimated from the samples.
success. With these premises, the natural choice would be to model the data with a Binomial distribution that describes the probability of $X$ successes in $n$ trials. Unlike the Binomial distribution (for which both $p$ and $\text{non-}p$ are known or can estimated conditionally on the predictors), distributions from the Poisson family assume no theoretical limit to the number of 'successes' that can occur (in our case, the number of non-self descriptions a participants could produce), and model the number of 'successes' occurring in a given interval. It seems, hence, that Poisson-like distributions would be a bad choice for our data. However, the Poisson distribution mimics the Binomial for very rare events (when probability $p < .10$). And this was exactly the case with the production of non-self descriptions in our experiment (where $p = .08$), as can also be seen in Figure A. This gave us further confidence in the approach. Furthermore, for reasons of over-dispersion explained above, we eventually settled for a (truncated) Negative Binomial instead of the (truncated) Poisson\textsuperscript{52}.

\footnote{Note that the Poisson is a special case of the Negative Binomial model when the over-dispersion parameter converges to zero [http://www.ats.ucla.edu/stat/stata/seminars/count_presentation/count.htm].}
Appendix T

Experiment 9: Justifications for using a two-stage hurdle approach

Data in the solo and the crowd conditions contained 20% of zero counts. Figure B reports the density plot of the two empirical distributions with superimposed theoretical Poisson, Binomial and Negative Binomial densities. It is difficult to establish whether 20% (rather than 40% or 60%) of zero counts constitute excessive zeros (cf. Zuur et al., 2012). Some authors have suggested that, in many situations, a Negative Binomial can capture the many zeros without resorting to hurdle (or zero-inflated) models (Warton, 2005). Figure B shows indeed that a Negative Binomial approximated the data better than the Poisson or the Binomial. For this reason, we decided to apply both a hurdle model and a Negative Binomial GLM to our data. Note that our design was between-subjects so there was no need to include any random intercept for subjects.

To decide between the two models we compared their predictions against the data. Figure C shows the mean predicted probability for each count value generated by the two models against the observed probability. For both condition, the Hurdle model was a better fit, also confirmed by AIC comparison (solo condition: hurdle AIC 468.45, Negative Binomial AIC 478.17; crowd condition: hurdle AIC 472.65, Negative Binomial AIC 484.38) and so it is the one we decided to report and discuss.
Figure B Density plots of the empirical (blue bars) and the Poisson (blue line), Negative Binomial (green line) and Binomial (violet) theoretical distributions for the Solo (upper panel) and Crowd (lower panel) data. The distribution parameters for the Poisson (solo \( \lambda = 3.03 \), crowd \( \lambda = 3.17 \)), the Negative Binomial (solo size = 0.86 and mu = 3.03, crowd size = 0.92 and mu = 3.17) and the Binomial (solo \( n = 36 \) and \( p = 0.08 \); crowd \( n = 36 \) and \( p = 0.09 \)) distributions were estimated from the samples.
Figure C: Density plots of the empirical distributions (blue bars) with predicted probabilities from the hurdle (red line) model and the negative binomial GLM (green line).