
THE UNIVERSITY
*of* EDINBURGH

# Mechanisms of Variation in Language Use: Insights from Lexical Entrainment

Anita Tobar Henríquez

PhD in Psychology
University of Edinburgh
2020

# 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.

Anita Tobar Henríquez

31st March 2020

# Lay Summary

Language use is variable, such that we can call the same object *umbrella* or *brolly*. But what determines this variability? This thesis aimed at better understanding what determines how people vary their referential expressions, both from person to person (*individual differences*) and from situation to situation (*situational factors*). In particular, we investigated the mechanisms underlying lexical entrainment, or the tendency for a speaker to reuse the same word as their interlocutor (e.g., using *brolly* after your partner used *brolly*; Brennan & Clark, 1996). Interestingly, speakers entrain to their partner's word even when that means using a word that is not their speech community's norm (e.g., using *brolly* instead of *umbrella* in British English; Branigan et al., 2011).

There are two main theoretical explanations of how lexical entrainment works. Unmediated accounts explain it as the result of lexical processing that is independent of speakers' beliefs, in that a partner's use of *brolly* made its lexical representation accessible from memory, thus enhancing its retrieval and reuse (Pickering & Garrod, 2004). In contrast, mediated accounts argue that lexical entrainment results from lexical processing that *is* mediated by beliefs, whether it is to enhance mutual comprehension (audience design; Clark, 1996) or to enhance social affiliation with an interlocutor (van Baaren et al., 2003). Critically, lexical retrieval, audience design, and social affiliation may vary both across individuals and situations, positioning lexical entrainment as a good candidate to investigate how individual differences and situational factors influence the variability of referential choices. Thus, in four studies, we investigated individual, interpersonal, and community-level influences on lexical entrainment.

Study 1 examined the extent to which lexical entrainment can reliably reflect individuals' variation in their propensity to reuse an interlocutor's word. We did this by testing the ability of an online web-based lexical entrainment task to elicit the same score for each individual across two sessions. Experimental items were objects

with both a favoured name (*umbrella*) and a disfavoured but acceptable name (*brolly*), and we measured lexical entrainment as participants' reuse of their partner's disfavoured name (i.e., using *brolly* after their partner used *brolly*). Across two studies, we found that the task consistently elicited the same scores for each individual, both across sessions separated by minutes and sessions separated by a week, clearly suggesting that the tendency to reuse a partner's lexical choice is stable within individuals and thus suggesting that individuals' tendency to lexically entrain can be underlain by stable individual traits.

Study 2 thus used this lexical entrainment task to examine whether entrainment was predicted by individual differences in schizotypy and age, which both correlate with individual differences in potential lexical entrainment mechanisms (i.e., lexical retrieval, audience design, and social affiliation skills). Although entrainment was not predicted by schizotypy, it was positively predicted by age, suggesting that at least some mechanisms of lexical entrainment may undergo changes across the lifespan.

Study 3 investigated whether lexical entrainment involved social affiliation goals. Based on previous evidence that ostracism increases social affiliative tendencies as a compensatory strategy to recover social acceptance, we examined (i) the effects of ostracism on lexical entrainment, (ii) whether such effects were targeted to repairing a particular social relationship or increasing affiliation more generally, and (iii) whether ostracism effects were moderated by personality. Across two experiments, we found that lexical entrainment was positively affected by ostracism, but ostracised participants were as likely to entrain to a partner who had ostracised them as to a new partner; importantly, post-ostracism lexical entrainment was predicted by individual differences in neuroticism. Critically, these results suggest a social affiliation component to lexical entrainment.

Study 4 moved beyond looking only at lexical entrainment with a particular partner, to examine how lexical entrainment can help speakers learn speech communities' lexical preferences (e.g., that *umbrella* is a favoured name and *brolly* a disfavoured name in the United Kingdom). Across three experiments, we found that participants generalised names across two partners depending on their community

membership. These results suggest that social information is encoded during lexical processing, thus enabling the creation of community-level knowledge from single linguistic encounters.

In sum, this research has important implications for understanding the variability of lexical choices. First, our individual differences findings indicate that lexical entrainment has potential to reveal what drives individual variation in how speakers make referential choices. Second, our group-comparisons findings indicate that the mechanisms of language processing are not encapsulated within the language system, thus suggesting that speakers vary their referential expressions based not only on linguistic processing but also on their beliefs and social dispositions. Taken together, these results suggest that the variability of language use is shaped by individual, interpersonal, and community-level influences, and by their interplay.

# Abstract

Language use is variable, such as we can call the same object *umbrella* or *brolly*. But what determines this variability? This thesis aimed at better understanding how speakers vary the words they use during dialogue, by investigating the mechanisms underlying lexical entrainment, or the tendency for a speaker to reuse a word that their partner has used before (e.g., using *brolly* after your partner used *brolly*; Brennan & Clark, 1996). Interestingly, speakers entrain to their partner's word even when that means using a word that is not their speech community's preference (e.g., using *brolly* instead of *umbrella* in British English; Branigan et al., 2011).

There are two main accounts of how lexical entrainment works. Unmediated accounts explain it as the result of priming effects, in that a partner's use of *brolly* made its lexical representation accessible from memory, thus enhancing its retrieval and reuse (Pickering & Garrod, 2004). In contrast, mediated accounts feature speakers' beliefs in entrainment, with some such accounts suggesting that entrainment is aimed to enhance mutual comprehension (Clark, 1996) and others arguing that it is aimed to enhance our social affiliation (van Baaren et al., 2003). Critically, lexical retrieval, audience design, and social affiliation may vary both across situations and individuals, positioning lexical entrainment as a good candidate to inform how individual, interpersonal, and community-level influences affect language use. Thus, in four studies, we investigated individual, interpersonal, and community-level influences in lexical entrainment.

The first research study examined the test-retest reliability of an online, web-based lexical entrainment task. Experimental items were objects with both a favoured name (*umbrella*) and a disfavoured but acceptable name (*brolly*), and we measured lexical entrainment as participants' reuse of their partner's disfavoured name (i.e., using *brolly* after their partner used *brolly*). Across two studies, we found that the task reliably elicited entrainment at the individual level both in the short- and the long-term, clearly suggesting that the tendency to reuse a partner's lexical choice is

stable within individuals and thus suggesting that individuals' tendency to lexically entrain can be underlain by stable individual traits..

The second study thus used this lexical entrainment task to test whether entrainment was predicted by schizotypy and age, which both correlate with individual differences in lexical retrieval, audience design, and social affiliation skills. Although entrainment was not predicted by schizotypy, it was positively predicted by age, suggesting that at least some mechanisms of lexical entrainment may undergo changes across the lifespan.

The third research project investigated the causal relationship between social affiliation and lexical entrainment. Based on previous evidence that ostracism increases social affiliation as a compensatory strategy to recover social acceptance, we examined (i) the effects of ostracism on lexical entrainment, (ii) whether such effects were targeted to repairing a particular social relationship or increasing affiliation more generally, and (iii) whether ostracism effects were moderated by personality. Across two experiments, we found that lexical entrainment was positively affected by ostracism, but ostracised participants were as likely to entrain to a partner who had ostracised them as to a new partner; moreover, post-ostracism lexical entrainment was predicted by individual differences in neuroticism. Critically, these results suggest a social affiliation component to lexical entrainment.

The fourth research project moved beyond looking only at lexical entrainment with a particular partner, to examine how speakers extrapolate community-level lexical knowledge from single linguistic encounters. Across three experiments, we found that participants generalised names across two partners depending on their community membership. These results suggest that social information is encoded during lexical processing, thus enabling the creation of community-level knowledge from single linguistic encounters.

In sum, this research has important implications for understanding the variability of lexical choices. First, our individual differences findings indicate that lexical entrainment has potential to reveal what drives individual variation in how speakers make referential choices. Second, our group-comparisons findings indicate that the mechanisms of language processing are not encapsulated within the language

system, thus suggesting that speakers vary their referential expressions based not only on linguistic processing but also on their beliefs and social dispositions. Taken together, these results suggest that the variability of language use is shaped by individual, interpersonal, and community-level influences, and by their interplay.

# Acknowledgements

I gratefully acknowledge the financial support I received from CONICYT, and I wish to thank every single participant who took part in my studies.

I would like to thank my examiners, Martin Pickering, Dale Barr, and Susan Brennan, for their thorough and encouraging feedback.

I express my deep gratitude to my supervisors, Holly Branigan and Hugh Rabagliati. Their guidance and support turned a process that I expected to be soul-draining into a safe space for learning and growth. Working with them has transformed my idea of what doing research should feel like, and I thank them for that sincerely.

I wish to thank my parents, Clementina and Luis, and my brother, Felipe. For supporting me in doing what I love the most, for having filled my childhood with heartwarming memories that I carry everywhere, and for giving me strength throughout this process.

I thank my dearest friends, for putting up with my obsession with language and conversation since we were little. Thank you guys for showing me what open-hearted communication can bring into one's life.

Finally, I want to thank Daniel, for being the person I trust the most in the whole world, and for encouraging me to cut down on perfectionism to give way to self-appreciation and kindness.

# Contents

Declaration....1
Lay Summary....2
Abstract....5
Acknowledgements....8
1. Introduction....11
1.1. What is this thesis about?....11
1.2. How is this thesis structured?....12
2. Mechanisms of lexical entrainment: A review....15
2.1. Introduction....15
2.2. The study of language use and lexical entrainment accounts....17
2.3. What could lexical entrainment tell us about the variability of language use? ....35
2.4. An empirical investigation of variation in language use through the lens of lexical entrainment....38
3. Lexical entrainment reflects a stable individual trait....40
3.0. Abstract....42
3.1. Introduction....43
3.2. Materials' creation: Norming tasks....50
3.3. Study 1: Short-term reliability....53
3.4. Study 2: Over-a-week reliability....60
3.5. Follow-up Analyses....64
3.6. Discussion....71
3.7. References....77
4. Individual differences in lexical entrainment....82
4.1. Introduction....82
4.2. Method....92
4.3. Results....96
4.4. Follow-up Analysis....101
4.5. Discussion....102
5. Lexical entrainment as an affiliative behaviour....107
5.1. Introduction....107
5.2. Experiment 1: Does lexical entrainment involve an affiliative component?.117
5.3. Experiment 2: A replication....133
5.4. Follow-up Analyses....140
5.5. Discussion....147
6. Speakers extrapolate community-level knowledge from individual linguistic encounters....154
6.1. Introduction....154
6.2. Experiment 1: Do speakers extrapolate community lexical preferences from individual linguistic encounters?....161
6.3. Experiment 2: Do the effects of partner's community membership on speakers' lexical expressions generalize across speech communities?....175

6.4. Experiment 3: Are community effects confounded with linguistic status?…182
6.5. Discussion.....187
7. Conclusions.....196
7.1. Introduction.....196
7.2. Overview of studies and findings.....197
7.3. Implications for accounts of lexical entrainment and future research.....201
7.4. Implications for the variability of language use.....206
7.5. Conclusions.....208
8. References.....209

Chapter 1

# 1. Introduction

## 1.1. What is this thesis about?

As language users, we know that how people put ideas into words varies widely across communicative situations, and also from person to person. This fundamental variability of language use shows up even in everyday references to commonplace objects; for example, British people can call the same thing an *umbrella* (a favoured or high-frequency name) or a *brolly* (a disfavoured or low-frequency name). However, we know very little about which cognitive mechanisms make this possible.

In order to understand the factors that affect variation in language use, the present thesis examined the phenomenon of lexical entrainment, or the tendency for a speaker to reuse the same words as their interlocutor has used before (**lexical entrainment**; e.g., using *brolly* after your interlocutor used *brolly*; Brennan & Clark, 1996). Interestingly, lexical entrainment occurs even when it means adopting an expression that speakers would not normally tend to use (e.g., using *brolly* instead of *umbrella*; Branigan, et al., 2011). Several theories have aimed to explain which cognitive processes underlie this phenomenon. *Unmediated accounts* of entrainment conceptualise this behaviour as the result of recent lexical processing that *is not* mediated by speakers' beliefs (e.g., Pickering & Garrod, 2004). In contrast, *mediated accounts* suggest that the phenomenon *is* mediated by beliefs, either as a way to enhance mutual comprehension during dialogue (i.e., audience design; Clark, 1996) or to pursue social goals (i.e., social affiliation; van Baaren et al., 2003). Interestingly, putative lexical entrainment mechanisms - lexical retrieval, audience

design, and social affiliation - can vary across both individuals and social situations. Taken together with the fact that lexical entrainment occurs even when that means using a disfavoured name in speakers' speech community (e.g., Branigan et al., 2011) the phenomenon offers a unique opportunity to understand how individual, interpersonal, and community-level factors inform and constrain speakers' referential choices during dialogue.

This thesis is an empirical attempt to further our understanding of the variability of language use, by examining three potential drivers: Individual differences, interpersonal factors, and community-level linguistic influences. In four studies, we address questions such as whether and how the tendency to entrain varies across speakers, whether and how speakers' beliefs about the social situation influence their propensity to reuse their partner's words, and how speakers generalise their own referential choices across different partners.

## 1.2. How is this thesis structured?

Chapter 2 presents the theoretical framework for this empirical investigation. In this chapter, we review what is known and unknown about lexical entrainment, identifying in the process what research on lexical entrainment could tell us about the variability of language use.

Chapter 3 examines the extent to which lexical entrainment can reliably reflect individual differences in how speakers make referential choices. Lexical entrainment is highly reliable for group-level comparisons, and thus has promise as a tool to understand how situational factors influence language use by using group-level experimental manipulations. But if we want to understand how lexical entrainment varies from person to person, then we need to make sure that lexical entrainment tasks are able to reliably distinguish between individuals. In this chapter, we describe in detail the lexical entrainment task we used across all our studies (with occasional modifications that will be presented where relevant). Most importantly, in this chapter we examine the extent to which our lexical entrainment task reliably

captures individual variation in lexical entrainment, by testing the test-retest reliability of the task across two studies.

Chapter 4 presents a preliminary investigation of the mechanisms that underlie individual differences in lexical entrainment. Based on previous evidence that lexical retrieval, audience design, and social affiliation become compromised with healthy ageing and the presence of schizotypical traits, we examined how age and schizotypy traits affect the likelihood of entrainment across individuals.

Chapter 5 moves on to the study of situational factors in lexical entrainment, by examining whether the tendency to entrain involves a social affiliation component. Although the repetition of a partner's language use has been suggested to imply a social affiliation component, the actual causal relationship between affiliative behaviours and lexical entrainment has not been empirically tested. Based on previous evidence that ostracism increases social affiliative behaviours as a compensatory strategy to recover social acceptance, we examined the effects of ostracism on lexical entrainment. Importantly, we also looked at (i) whether ostracism effects on entrainment were targeted to repairing a particular social relationship (with the perpetrator of ostracism) or increasing affiliation more generally (with any social partner), and (ii) whether ostracism effects were moderated by individual differences in neuroticism, which correlates with affiliation dispositions.

Chapter 6 goes beyond looking only at lexical entrainment with a particular partner, to also examine how speakers extrapolate community-level lexical knowledge from single linguistic encounters. Most research on how audience design shapes language use has focused on which information is shared between two particular interlocutors, but how community-level influences affect the relationship between language use and audience design has not received much attention. Indeed, we do not know much about how speakers establish community-level linguistic knowledge in the first place. Chapter 6 addresses these issues. Across a series of experiments, we first test how speakers vary their lexical choices with a partner depending on that partner's choices and speech community, and we then test whether

speakers extrapolate their own referential choices with a first partner to a subsequent partner depending on their community membership.

Chapter 7 discusses the implications of this thesis for understanding lexical entrainment and the variability of language use. We first summarise the results of each study (Chapters 3 to 6), and then discuss the implications of these findings for theories of lexical entrainment, in particular, and for understanding the variability of language use, more generally.

Note: Each empirical chapter (i.e., Chapters 3 to 6) is written as a self-contained paper. For the sake of global coherence and to enhance exposition, some chapters are prefaced by a brief introductory note that explains how the chapter relates to the overarching narrative of this doctoral investigation.

Chapter 2

# 2. Mechanisms of lexical entrainment: A review

## 2.1. Introduction

Language use is intrinsically variable: People express themselves differently across different social situations (Eckert & Rickford, 2001; Gregory & Carroll, 2018), and across their lifespans (e.g., Juncos-Rabadán, Facal, Rodríguez, & Pereiro, 2010; March, Wales, & Pattison, 2011). Language use also shows wide variation from individual to individual. Interestingly, this variability shows up even in references to commonplace objects, where speakers might variously refer to the same object as *umbrella*, *brolly* or even just *thing*. Thus, understanding what determines the variability of speakers' lexical choices can illuminate which mechanisms underlie the variability of language use more generally.

One possibility is that speakers vary their lexical choices depending on features of the communicative situation (*situational factors*), such as which words they have recently heard (Pickering & Garrod, 2004), which words they believe that their interlocutor prefers (Clark & Brennan, 1996), or whether they want to increase rapport with a particular interlocutor (e.g., Giles, Mulac, Bradac, & Johnson, 1987; Van Baaren, Holland, Steenaert, van Knippenberg, 2003). Another possibility is that this variability is also moderated by individual differences in how speakers' minds process linguistic and social information. After all, linguistic communication during interaction involves quite complicated processes, such as accessing appropriate

referential expressions from memory and monitoring relevant contextual cues to achieve mutual comprehension, and how the mind handles these processes is likely to vary from person to person (e.g., Brown-Schmidt, 2009). Moreover, it is possible that variation in language use is moderated by community-level influences, such as which are the most frequently used labels in the speakers' linguistic community and their knowledge of other communities' lexical preferences (e.g., Isaacs & Clark, 1987).

Critically, it is uncontroversial in psycholinguistics that speakers' lexical choices are influenced by their interlocutor's choices. For example, they tend to reuse their partner's words to refer to the same objects (i.e., ***lexical entrainment***, e.g., using *brolly* after their partner used *brolly*; Branigan, Pickering, Pearson, McLean, & Brown, 2011; Clark and Wilkes-Gibbs, 1986; Garrod & Anderson, 1987; Brennan & Clark, 1996).

However, the processes that underlie lexical entrainment are still unclear. *Unmediated accounts* have explained the phenomenon as a result of recent lexical processing that is not mediated by speakers' beliefs, in that a partner's use of *brolly* made the word accessible from memory, thus enhancing lexical retrieval and reuse of the word (Pickering & Garrod, 2004). *Mediated accounts*, in contrast, have argued that lexical entrainment is mediated by beliefs, either as a way to enhance mutual comprehension during dialogue (i.e., audience design; Brennan & Clark, 1996) or to achieve social goals with a partner (i.e., social affiliation; Giles et al., 1987; van Baaren et al., 2003).

There are mixed results regarding whether lexical entrainment emerges from mediated or unmediated processing, and none of these views fully explains which cognitive mechanisms underlie the phenomenon. However, what it is clear from lexical entrainment literature is that its potential underlying mechanisms - lexical retrieval, audience design, and social affiliation - can vary across different individuals and different social situations. Moreover, if lexical entrainment is affected by audience design and social affiliation, then the tendency to lexically entrain is most likely to vary depending on the interlocutors' speech communities. Thus the

phenomenon offers a unique opportunity to understand how individual, interactional, and community-level factors determine the variability of referential choices.

This chapter reviews what we know and do not know about how lexical entrainment works, aiming to identify how furthering our understanding of lexical entrainment can illuminate which individual, situational, and community-level factors inform and restrict speakers' lexical choices. In the first part of the chapter, we present theoretical accounts of lexical entrainment. We start by presenting unmediated accounts and then turn to mediated accounts, where we first concentrate on accounts featuring audience design, and then move on to accounts featuring social affiliation. In the second part of this chapter, we discuss how furthering our understanding of how lexical entrainment works can help us understand better which individual, interactional, and community-level factors affect the variability of speakers' referential expressions. It is worth noting this thesis does not aim to prove or disprove unmediated or mediated theories of entrainment, but rather to understand better the mechanisms underlying the phenomenon.

## 2.2. The study of language use and lexical entrainment accounts

The study of language use during dialogue started drawing attention from psycholinguists only a few decades ago, when we moved from understanding language use as a decontextualized phenomenon to conceptualising it as a social action. The main advocate of the experimental study of dialogue is Herbert Clark, who proposed that language use is affected by speakers' communicative intentions, and whose seminal research on dialogue influenced current accounts of language use during dialogue.

In a seminal study, Clark and Wilkes-Gibbs (1986) proposed that a major factor for successful communication during dialogue is the minimization of collaborative effort, implying that interlocutors formulate their utterances so that they do not need to spend unnecessary time or effort in ensuring mutual intelligibility. Consistent with this idea, and importantly for theories of lexical entrainment, Garrod and Anderson (1987) suggested that speakers formulate utterances according to the

same principles of interpretation as those needed to interpret utterances recently used by their addressees (*output/input co-ordination* principle). Under this logic, interlocutors should be locally consistent with each other, and they should use the same linguistic forms to refer to the same state of affairs, achieving a mutually satisfactory description of the situation under discussion with the minimum effort. For example, if your partner describes an action using a passive construction (e.g., *The chocolate was eaten*) then you should also use a passive construction to describe an action in a subsequent opportunity (e.g., *The dishes were washed up*).

However, Clark and colleagues' seminal work did not explain which basic processing mechanisms made this co-ordination principle possible, which encouraged the diversification of the psycholinguistics of language use, leading to the development of the unmediated and mediated accounts we discuss below.

But before discussing mediated and unmediated theories any further, we need to make an important terminological consideration. Ironically, the speakers' reuse of their partner's words has been variously referred to as *accommodation*, *convergence*, *entrainment*, and *alignment*. But these terms have been used with slightly different meanings that are worth clarifying. *Accommodation* is generally used to refer to the general adaptation of verbal behaviour in all linguistic levels, ranging from pronunciation to linguistic style (e.g., Giles et al., 1987); importantly, this term has been used to refer to both adaptations that make speakers' language more similar to their interlocutor's, or more different from their interlocutor's. *Convergence* has been used to talk about interlocutors' general tendency to make their language use similar to each other's at multiple levels of linguistic structure (e.g., including pronunciation, pitch, or intonation; Pardo, Urmanche, Wilman, & Wiener, 2017). *Entrainment* has been used to refer to speakers' tendency to reuse a referential expression that has been recently used during interaction, by either their partner or themselves (e.g., using *brolly* after having recently used or heard *brolly* during dialogue; Brennan & Clark, 1996). Finally, *alignment* has been described as the simultaneous activation of linguistic mental representations, e.g., when both interlocutors access the lexical representation of *brolly* simultaneously, due to simultaneous recent processing (e.g., Pickering & Garrod, 2004). Problematically, researchers have used *alignment* to refer

to both the alignment of mental representations and the speaker's reuse of a partner's language use indiscriminately, which may cause confusion as to whether researchers assume an unmediated or a mediated approach to the phenomenon (e.g., Kaschak & Glenberg, 2004; Branigan, Pickering, Pearson & McLean, 2010; Xu & Reitter, 2015; Doyle & Frank, 2016).

In this review (and thesis), we conceptualise lexical entrainment as a linguistic behaviour and, in particular, as the speaker's overt use of a word that *their partner* has used before. As we aim to understand which mechanisms underlie this linguistic behaviour, we do not assume an unmediated or a mediated perspective a priori. Then, throughout this chapter we will use *entrainment*[1] to refer to a speaker's overt use of a partner's linguistic choice, *alignment* to refer to the simultaneous activation of interlocutors' linguistic representations, *convergence* to refer to interlocutors' tendency to make their language use more similar to each other, and *accommodation* to refer to speakers' general tendency to adapt their language use during interaction.

### 2.2.1. Unmediated accounts: Lexical entrainment as a result of how we process language

One prominent unmediated account of language use is the *Interactive Alignment Account of Dialogue* (Pickering and Garrod, 2004). Consistent with Clark's foundational work, this account assumes that production and comprehension become coupled during conversation, but in contrast to Clark's assumptions (discussed below), it implies that entrainment is driven by unmediated linguistic processing that is independent of speakers' beliefs and social dispositions. In particular, Pickering & Garrod suggest that speakers' entrained linguistic behaviours

[1]It is worth noting that Brennan & Clark (1996) defined *entrainment* as a process, where a speaker's proposed referring expression is ratified by their interlocutor's reuse of that word or by any signs that they had understood the term. In this thesis, however, we use *entrainment* to refer to a speaker's behaviour and not a process deployed by a dyad of interlocutors. In particular, we define *entrainment* as the overt reuse of a partner's term (i.e., using *brolly* after a partner used *brolly*).

leads to alignment of each other's linguistic representations, and vice versa, mainly due to linguistic priming (see Neely, 1976).

As shown in Figure 1, under this account different levels of linguistic representation are connected to each other, so that the activation of one of them may lead to the activation of others. For instance, when a speaker uses *brolly* during dialogue, the interlocutors' simultaneous processing of the word (i.e., the speaker's production of *brolly* and the addressee's comprehension of *brolly*) leads to the mutual activation of each other's lexical representation of the word (BROLLY). This in turn leads to the alignment of the appropriate semantic representations (what *brolly* means), to then reach the situation model's level, achieving mutual comprehension.

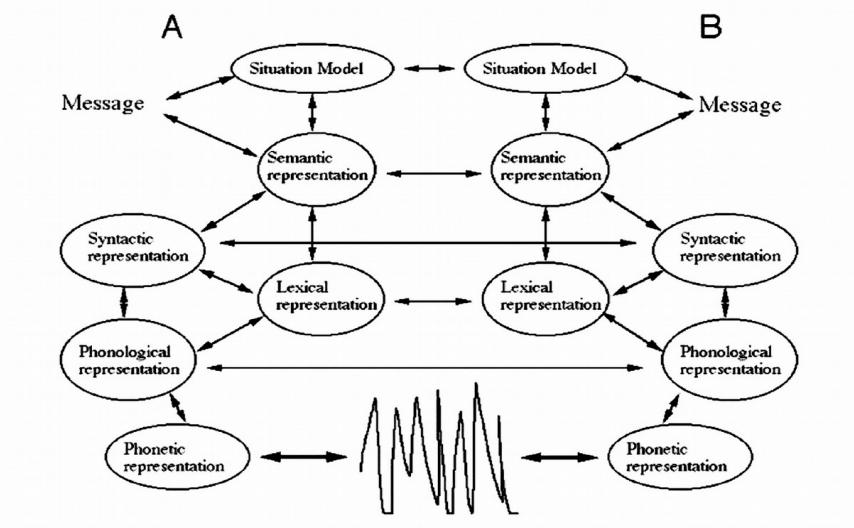


**Figure 2.1.** Interactive Alignment model (Pickering & Garrod, 2004, p. 8). A and B represent two interlocutors who achieve mutual understanding by means of the coordination of their linguistic behaviours. Vertical and diagonal arrows represent connections between a speaker's levels of linguistic representation. Horizontal arrows represent connections between interlocutors' same levels of linguistic representation. At the bottom, the line-drawing of a pressure oscillation represents interlocutors' language use.

Under this logic, lexical entrainment is explained as a consequence of lexical priming effects. Specifically, speakers entrained to a partner's use of *brolly* because the recent processing of the word makes its lexical representation accessible from memory, thus enhancing its retrieval and reuse.

Several studies support an unmediated component to lexical entrainment. In particular, classical priming studies have shown that having recently processed a word makes it more accessible from memory. For instance, you will respond faster to the word *chocolate* after having recently processed *chocolate* than after having processed another word, and you will even respond faster to the word *chocolate* after having processed a semantically related word, like *brownie*, than after having processed a semantically unrelated word, like *computer* (Neely, 1976; Meyer, 1996). Moreover, research on syntactic priming has shown that recently processed syntactic forms affect subsequent production at the syntactic level. For example, you will be more likely to use a passive construction after your partner used a passive construction than after they used an active construction (e.g., Bock, 1986).

Critically, there is also evidence that priming effects percolate across different levels of linguistic representation. In particular, Branigan, Pickering and Cleland (2000) found that participants not only tended to reuse their partner's syntactic choices, but also that this tendency increased when they were asked to repeat the same verb as their interlocutor versus using a new verb (see also Rowland, Chang, Ambridge, Pine, & Lieven, 2012; Segaert, Kempen, Petersson, Hagoort, 2013; Traxler, Tooley, Pickering, 2014). These results importantly indicate that entrainment at the lexical level (i.e., repeating the verb) enhanced entrainment at the syntactic level (i.e., repeating the syntactic structure), which in turn suggests that priming effects at one particular level of linguistic structure can percolate other levels of representation.

Moreover, speakers reuse recently processed referential expressions even when such expressions are not the most appropriate for the context (see Schiffer, 1972; Grice, 1975), suggesting an unmediated component to the tendency to reuse recently encountered terms. For example, Brennan & Clark (1996) had pairs of speakers participate in two consecutive sessions of an interactive referential task,

where they had to describe to each other a set of cards depicting objects, e.g., a men's pennyloafer (see also Fussell & Krauss, 1991). In the first session, the target objects were displayed along with several other tokens from the same category, e.g., participants saw several men's pennyloafers. In the second session, the target objects were the only members of their category; e.g., participants saw only one candle. Importantly, participants not only tended to use subordinate names (e.g., *the men's dress shoe*) to name objects in the first session, but they maintained those over-specified referential expressions in the second session too, where using a subordinate name was no longer necessary for their partner to identify the target.

Furthermore, Branigan and colleagues have consistently shown recent lexical processing effects in speakers' tendency to lexically entrain. For example, Branigan et al. (2011) developed a lexical entrainment task where participants collaborate with a confederate to match and name pictures. The experimental targets are pictures of objects that can be named with both a favoured label and a disfavoured label in participants' speech community (e.g., *umbrella* versus *brolly*, in British English). These materials have been pre-tested to ensure that participants rarely use the disfavoured label spontaneously, but still consider it an acceptable name for the object. In the main task, participants always name the experimental targets after the confederate, and lexical entrainment is then measured as the proportion of trials on which participants use the same disfavoured label used by the confederate.

Branigan and colleagues found that individuals used a disfavoured label (e.g., *brolly*) more often after their partner had used that label than after their partner had used the favoured label (e.g., *umbrella*) to name the target, or compared to the favoured label's baseline frequency of use in spontaneous naming (Branigan et al., 2011, 2016; Hopkins et al., 2017). Moreover, they showed that speakers' propensity to reuse a partner's disfavoured label was importantly affected by the lag between the prime trial, i.e., where participants matched the target with its disfavoured label, and its corresponding naming trial, i.e., where participants named the previously matched target. Critically, the longer the prime-target lag was, the less likely participants were to reuse their partner's disfavoured label (Branigan et al., 2011). Taken together, these results support that recent lexical processing importantly influences speakers'

referential expressions during dialogue, suggesting an unmediated component to lexical entrainment.

In sum, unmediated accounts of language use explain lexical entrainment as a consequence of language processing mechanisms, and thus suggest that the phenomenon is not mediated by audience design or social goals. In particular, they propose that lexical entrainment is mainly driven by lexical priming, so that speakers reuse their partner's choice because the recent processing of that word makes its lexical representation accessible from memory, enhancing its retrieval and use. There is convincing evidence in favour of an unmediated component to lexical entrainment, which is mostly based on experimental group-comparisons studies. However, other accounts of language processing argue that this evidence is not enough to suggest that entrainment is mainly driven by unmediated, priming effects. Importantly, the unmediated approach thus suggests that lexical entrainment might vary from person to person depending on individual differences in priming effects and lexical retrieval.

### 2.2.2. Mediated accounts: Lexical entrainment is influenced by audience design and social affiliation goals

Mediated accounts of dialogue focus on the roles of speakers' beliefs in how they use language. Some such accounts focus mainly on the role of *audience design* in language use, i.e. the speaker's taking into account of what the addressee is able to understand or not (e.g., Clark, 1996; see also Bell, 1984). Other accounts focus on the role of social affiliation and other social factors in language use (e.g., van Baaren et al., 2003; Giles et al., 1987).

#### 2.2.2.1. Lexical entrainment and audience design

A key term to understand audience design accounts is *common ground*, this is information that interlocutors assume to be mutually shared (Clark & Marshall, 1986; Clark, 1996). Two sorts of common ground have been contrasted (see Clark, 1996): non-linguistic common ground (i.e., world knowledge) and linguistic common

ground (i.e., lexicon, grammar, and pragmatic expectations of language use). In particular, linguistic common ground can be considered as personal or communal (Clark & Marshall, 1986). Linguistic personal common ground describes linguistic information that interlocutors take to be mutually known based on their joint personal experience (copresence heuristics; see Clark & Marshall, 1986, p. 34). For example, when you experience your conversational partner using *brolly,* the label becomes part of the personal linguistic common ground that you share with that partner. In contrast, linguistic communal common ground describes the linguistic knowledge, preferences, and expectations of language use that are shared by members of a specific social group (i.e., speech community; see Gumperz, 1968, 2009; Labov, 1972), and as such this kind of common ground does not necessarily depend on copresence heuristics. For example, it is linguistic communal common ground shared by British people that an umbrella can be named with both the (British) favoured label *umbrella* and the disfavoured (but acceptable) label *brolly*.

Audience design accounts argue that language is used to collaboratively establish and maintain linguistic personal common ground, thus determining many aspects of reference (e.g., Brennan, 1990; Brennan & Clark, 1996; Clark & Clark, 1979; Clark & Wilkes-Gibbs, 1986; Lockridge & Brennan, 2002). Under this logic, dialogue has been conceptualised as a type of joint action where interlocutors cooperate to achieve goals that depend on mutual understanding, and personal linguistic common ground becomes the relevant context against which this cooperation takes place.

Importantly, however, communal common ground can influence language use and mutual understanding too, since a speaker may need to adapt the words they use when interacting with people from other social groups or speech communities, who may not be familiar with the speaker's communal lexical preferences (e.g., that *umbrella* is favoured compared to *brolly* in the United Kingdom; Lewis, 1969; Clark, 1996; E. Clark, 1990, 2007). For instance, Isaacs & Clark (1987) found that New Yorkers named the Rockefeller Centre, an iconic New York landmark, using its proper name when their conversational partner was also from New York, but described it as '*the building with the flags outside*' when conversing with someone

from outside the tri-state area, who had failed to understand the proper name (see also Bortfeld & Brennan, 1997).

These findings show that language use can be importantly influenced by the extent to which interlocutors share communal common ground - but how speakers establish community-level linguistic knowledge is still an open question in the psycholinguistics of dialogue. Instead, audience design accounts of language use have focused mainly on how speakers establish personal common ground, and how this process affects language use and comprehension, suggesting that mutual comprehension is achieved only when interlocutors monitor and use personal common ground effectively.

But what does 'using common ground effectively' mean? Clark and Wilkes-Gibbs (1986) claimed that minimization of collaborative effort drives language use in dialogue. They assume that, for each reference, interlocutors try to reach the mutual belief that the addressee has identified the intended referent correctly. In the simplest case, the speaker presents a noun phrase (e.g. *brolly*) and the addressee accepts it by continuing the conversation. In other cases, the addressee may not be able to follow what the speaker means (e.g. if your partner would not know the word *brolly*), forcing the interlocutor to repair, expand or replace the recently used referential expression (e.g., by using the more frequent name *umbrella* or a more descriptive expression, such as *that thingy you use to cover from the rain*). These actions depends on co-presence heuristics and ends when both interlocutors accept that the description has finally arrived at its appropriate form; this is when the addressee gives signs of successful comprehension. Thus, making optimal use of common ground means, under classic audience design accounts, that both interlocutors take into account shared knowledge in order to make sure that the addressee always understands what the speaker means.

There are plenty of studies supporting an audience design component in lexical entrainment. In particular, Brennan and Clark (1996) demonstrated that speakers' tendency to use recently processed referential expressions is better accounted for by linguistic common ground between interlocutors, than by informativeness or recency alone. Although they found that participants reused over-

specified referential expressions when not necessary (i.e., *the man's dress shoe* when there was only one shoe), participants were still more likely to reuse those expression when they were in personal common ground versus when they were not.

Let us recall that in Brennan & Clark (1996)'s experiments participants engaged in two sessions. In Experiment 3, they manipulated whether or not the referential expressions used in the first session were in personal common ground during the second session: Participants either interacted with the same partner throughout the task or swapped partners between sessions. Participants maintained referential expressions more often when interacting with the same partner in both sessions than when switching partners in between, suggesting that speakers' vary their choices based on linguistic personal common ground. Critically, this finding has been repeatedly replicated (e.g., Horton & Gerrig, 2002, 2005), and it is consistent with evidence that comprehenders experience more difficulty understanding a partner's new label when that same partner has previously used another label versus when they have not (e.g., Metzing & Brennan, 2003; for a review see Kronmüller & Barr, 2015).

Moreover, in their lexical entrainment experiments, Branigan et al. (2011) manipulated participants' beliefs about whether their 'partner' was a computer or a human, and if the partner was a computer, whether it was more or less capable. Participants not only entrained more often to computer-partners than to human-partners, but also entrained more often to 'less capable' computer-partners than to 'more capable' computer-partners. These results suggest that participants entrained to a greater extent when they were less confident about their partner's understanding of the favoured term, which in turn suggests that lexical entrainment is sensitive to audience design.

Branigan et al. (2011)'s results also point out the importance of understanding better the interplay between communal and personal common ground in lexical entrainment. Given differences in their amount of previous linguistic experience, people have weaker models of communal lexical preferences for English-speaking computers than for English-speaking humans (Branigan, Pickering, Pearson & McLean, 2010). Thus, they might have interpreted computers' use of disfavoured

labels as evidence that those labels were the computer-community's preferences, and then assumed that computers might not understand the English favoured label, leading to an increased likelihood of lexical entrainment with computers versus humans.

Taken together, these results have been interpreted as evidence that lexical entrainment may involve an audience design component, but exactly how such audience design occurs is still under debate. Clark and Marshall (1978, 1981) argued that during audience design language users relied upon a set of simple co-presence heuristics that direct attention towards types of evidence that support inferences about what is known by both interlocutors. This process was not thought to require complex meta-representations of mutual knowledge, such as 'I believe that you know the word *brolly* (and that is why I think this particular sentence will be understandable to you)', or the interaction between linguistic processing and general cognitive processing, such as memory or executive functions. Instead, co-presence information would be encoded via special-purpose memory representations, in the form of reference diaries, i.e., episodic memory relevant to personal common ground, and of reference encyclopaedias, i.e. semantic memory relevant to communal common ground. Such memory representations about the co-presence of the speaker, the addressee, and the referent, provide a basis upon which language users may infer common ground in order to resolve and produce reference.

It is mainly regarding this particular consideration that Clark and Marshall's proposal has been criticized, motivating numerous theoretical accounts of how common ground is used, with most of them focusing primarily on language comprehension processes (for an overview, see Brennan & Hanna, 2009). Currently, the debate centres around how and when the language processing system uses common ground. In attempts to answer these questions, three prominent models of on-line perspective-taking have emerged: the Perspective-Adjustment model, the Anticipation-Integration model, and the Constraint-Based model.

The Perspective-Adjustment model starts from the assumption that audience design is resource intensive, and argues that most of the time it is unnecessary because the interlocutors' perspectives tend to overlap, and thus they can rely on

working memory to interpret sentences (Keysar, Barr, Balin, Paek, 1998; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003; Keysar, 2007). The model claims that the initial interpretation of a referring expression is egocentric, and that a second monitoring process checks for violations of common ground, adjusting the initial interpretation as needed (Keysar, et al., 1998).

These claims are consistent with evidence that people's estimates of others' knowledge are biased in the direction of their own knowledge. For example, Fussell and Krauss (1992) found that participants who knew the name of rare objects assumed that that particular object's name would be known by other individuals too (see also Fussell and Krauss, 1991, and Nickerson, Baddeley, & Freeman, 1987). More strikingly, Horton & Keysar (1996) found evidence to suggest that speakers use common ground to tailor their sentences only when they are not under time pressure, which they interpreted as evidence that speakers do not engage in audience design in the initial planning of utterances, but instead monitor those plans for violations of common ground. This model is also supported by the finding that comprehenders often use information that is not available to the speaker – that is in *privileged ground* - to interpret their utterances (Keysar, et al., 1998; Keysar, et al., 2000; Keysar, et al., 2003). For example, Keysar et al. (2003) tracked participants' gaze while they followed instructions to manipulate objects, such as '*Pick up the tape*', in contexts that included a cassette tape in common ground and a roll of Scotch tape in privileged ground. They found that participants were more likely to look at the privileged ground Scotch tape compared to a control condition where the privileged ground item was a battery, which indicates that language comprehension was influenced by egocentric processing.

The Anticipation-Integration model (Barr, 2008) argues that common ground helps establish pragmatic expectations prior to language processing. In line with the Perspective-Adjustment model, this model also assumes that common ground is not integrated with linguistic information during early language comprehension. Support for this view comes from evidence that listeners tend to create the pragmatic expectation that their partner will refer to an object in common ground, but during interpretation of referential expressions listeners tend to ignore which objects are

shared or privileged. For example, Barr (2008) recorded listeners' eye movements as they searched for the target of a speaker's referring expression in a display that also contained a phonological competitor (e.g., *bucket/buckle*). Listeners anticipated that the speaker would refer to something in common ground, i.e., before the onset of the partner's refererring expression, participants tended to look at objects that were shared with the speaker. However, they did not seem to take common ground into account during language comprehension itself: while processing the speaker's referential expression, participants experienced the same amount of interference from a competitor that only they could see compared to a matched competitor that was in common ground.

In contrast, the Constraint-Based model claims that audience design is one of many other constraints affecting language processing, and thus suggests that to assess if an individual was sensitive to audience design when interpreting a reference requires considering other sources of information that may have influenced linguistic processing (Nadig & Sedivy, 2002; Hanna, Tanenhaus, & Trueswell, 2003; Hanna & Tanenhaus, 2004; Heller, Grodner, & Tanenhaus, 2008; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Brown-Schmidt, 2009).

This model is supported by evidence that several sources of information affect language comprehension. For example, Hanna, et al. (2003) had participants interpret confederate's instructions, such as *Put the blue triangle on the red one*, in contexts where they could see one blue triangle and two red triangles. They manipulated whether both red triangles were in common ground, or if one was in common ground and the other was in the addressee's privileged ground. When both red triangles were in common ground, participants looked at both red triangles at similar rates, but when one of the red triangles was in privileged ground, participants looked more often at the common ground triangle within the first few hundred milliseconds of processing. However, there still was a lexical competition effect when one red triangle was in in privileged ground, and participants fixated the privileged ground competitor triangle more often when it was red compared to a condition in which it was yellow, suggesting that common ground competed with other processing demands. Moreover, previous work has suggested that the ability to

inhibit irrelevant contextual information determines the degree to which addressees successfully inhibit perspective-inappropriate interpretations of temporary referential ambiguities in their partner's speech, supporting that the use of common ground is moderated by cognitive demands (Brown-Schmidt, 2009).

Importantly, the idea that the use of common ground is restricted by processing demands is also supported by language production studies. For example, Wardlow Lane and Ferreira (2008) demonstrated the importance of general mechanisms of attention allocation during language production. They had participants describe targets (e.g., a candle) to naïve participants, and they manipulated whether competing targets (e.g., other candles varying in size) were in common ground or in privileged ground. Strikingly, they found that privileged objects influenced participants' referential expressions: Even when their partner could see only one candle, participants still used descriptions like *the big candle*. Importantly, they also found that speakers' referential expressions reflected use of privileged ground (i.e., information that was not available to their partner) especially when privileged objects were highly salient, supporting that audience design is only one of many competing constraints during language processing. Moreover, previous work suggests that older adults are more likely to use ambiguous descriptions for a target than their younger counterparts (e.g., saying *the spider* when the listener can see both a small and a big spider), and that this differential tendency correlates with individual differences in switching and inhibition skills, which supports that the use of common ground is restricted by cognitive load (Long, Horton, Rohde, & Sorace, 2018).

Consistent with the idea that language processing is determined by general cognitive processing and contra Clark and Marshall's (1978, 1981) idea of special-purpose memory representations supporting the use of common ground, Horton and Gerrig (2005, 2016) suggested that inferences about common ground relied on cue-driven retrieval processes found in global matching models of recognition memory (e.g., Ratcliff & McKoon, 1988). Cue-driven recognition models assume that input information functions as probes to memory, triggering a parallel search for any stored information that shares features with the probe cue. This search process is automatic,

and is referred to as *resonance* (Ratcliff, 1978), with *resonance strength* being the number of overlapping features between the stimulus cues and the target memory traces. For example, when resonance strength for a given stimulus (e.g., your partner's identity) reaches a certain activation threshold, related memories become accessible (e.g., the word *brolly*, and the memory that your partner used *brolly*), influencing subsequent processes (e.g., your choice to use *brolly*).

Under this logic, the assessment of what is mutually known emerges from a speaker's automatic recognition that some specific information can be treated as familiar within a particular context, thus affecting the likelihood of speakers to use particular forms of reference that are accessible in the moment. Although this search process is automatic, the overall activation of memories regarding a particular addressee will influence the likelihood that speakers may make strategic use of common ground during language production. For example, if you see an umbrella and remember that your partner has used *brolly* to name that object before, you may interpret their (previously stored) use of *brolly* as evidence of their understanding of the word, and may thus use *brolly* in a future naming opportunity to enhance communication with that partner.

These considerations are supported by evidence that speakers tailor their referential expressions based on common ground to the extent that they can access clear cues about 'who knows what'. For example, Horton & Gerrig (2005) had participants (acting as Directors) collaborate with two partners (acting as Matchers) to arrange sets of picture cards. In an orthogonal condition, each Matcher arranged cards from distinct card categories (e.g., Matcher A only saw dogs while Matcher B only saw fish); in an overlapping condition, each Matcher arranged unique cards from the same categories (e.g., Matcher A and Matcher B each saw different fish). In a subsequent round, Directors in the orthogonal condition elaborated more when describing cards that were new for a given Matcher, compared to Directors in the overlapping condition. These results suggest that the use of common ground will depend in many circumstances on the extent to which speakers have access to suitable memory representations.

In sum, audience design accounts of dialogue highlight the role of common ground in the variability of speakers' referential expressions, and thus explain lexical entrainment as a way to enhance communication with a partner. Such accounts do not assume that the use of common ground in language production implies the conscious and rational attribution of mental states. Regarding the cognitive mechanisms involved in lexical entrainment, classical audience design accounts suggest that the speakers' ability to track and use common ground depends on co-presence heuristics that rest upon language-specific diaries (Clark, 1996). Current audience design accounts, however, suggest that the use of common ground might be supported by general cognitive abilities, like cue-driven recognition and executive functions (Horton & Gerrig, 2002, 2005; Brown-Schmidt, 2009). Evidence for these accounts comes mostly from group-level comparisons, with only a few studies investigating what determines variability of referential expressions across individuals. Moreover, most audience design research has focused how speakers establish and use personal common ground to achieve mutual comprehension, but how they establish and use communal common ground has been largely neglected in the literature.

#### 2.2.2.2. Lexical entrainment and social affiliation

Under social affiliation accounts, lexical entrainment can be conceptualised as one of many other imitative behaviours. During social interaction, people not only entrain to their interlocutor's language use, but also to other non-linguistic behaviors, including body postures, gestures, and facial expressions (i.e., the 'chameleon effect'; see Chartrand & Lakin [2013] for a review). This tendency is often involuntary and unconscious, and it does not seem to be linked to the specific relationship between social partners. For example, Chartrand & Bargh (1999) found that participants engaged in more foot-shaking when interacting with a foot-shaking than a face-touching confederate, and engaged in more face-touching when interacting with a face-touching than a foot-shaking confederate. This differential tendency took place despite whether the participants knew the confederate, whether they were later able to recall the confederate's behaviour, and whether the confederate performed welcoming behaviours toward the participants.

What exactly drives behavioural mimicry does not have a unified answer, but it is uncontroversial that affiliation goals play a critical role. Previous studies suggest that people are more likely to imitate their partner's behaviours, like face-touching, when they are explicitly encouraged to get along with that partner compared to when they are not, and this increased affiliative mimicry occurs even when people are only primed with affiliative lexical labels, such as *affiliation*, *friend*, and *together* (e.g., Lakin & Chartrand, 2003; Leighton, Bird, Orsini, & Heyes, 2010). Moreover, previous studies suggest that experiencing ostracism, which can cause pervasive emotional pain, leads people to show increased behavioural mimicry as a way to recover social acceptance and belonging (Lakin, Chartrand, & Arkin, 2008). Based on this evidence, social psychology has suggested a bidirectional link between behavioural imitation and social affiliation, i.e., the 'perception-behaviour link' (Dijksterhuis & Bargh, 2001), which implies that people experience increased liking of partners who mimic their mannerisms, and tend to mimic partners with whom they want to affiliate more than those with whom they do not (Chartrand & Bargh, 1999; Stel & Vonk, 2010).

Based on these considerations, some social psychologists have suggested that the tendency to reuse a partner's language use could be understood as an expression of social affiliation (van Baaren et al., 2003; Chartrand, Maddux, & Lakin, 2005). Under a social affiliation account of lexical entrainment, speakers would reuse their partner's referential expressions to express affiliation and enhance social relationships, making interactions both effective and rewarding (Van Baaren et al., 2003; Reitter & Moore, 2014). For example, van Baaren and colleagues found that the repetition of a partner's language use increased that partner's pro-social behaviours in a surprisingly concrete manner. In particular, the authors had a waiter either mimic half their customers by repeating their order verbatim or not mimic them, and they then compared the amounts of tip left by mimicked versus non-mimicked costumers. They found that mimicked customers' tips were larger than non-mimicked customers' tips, suggesting that lexical entrainment increases partner's social affiliation dispositions.

These results and considerations are consistent with Communication Accommodation Theory (CAT), which assumes an unconscious affiliation motivation in speakers' tendency to accommodate (or adapt) their language use during interaction (Giles & Coupland, 1991; Giles et al., 1987). More specifically, this theory assumes that speakers' adaptation of their linguistic choices helps increase liking and decrease social distance between interlocutors, and argues that the degree of accommodation is moderated by multiple factors, from personality traits to in-group versus out-group identity (Natale, 1973; Giles, 1973).

CAT is supported by a vast number of studies. For example, previous work has suggested that individuals accommodate their vocal intensity (i.e., the loudness of their voice) to their partner's as a function of their self-reported need for social approval (Natale, 1973), and people tend to converge to a partner's pronunciation more often when the partner is from a high-status social group versus a low-status group (Gregory & Webster, 1996). Moreover, in intercultural settings, individuals from low-status social groups tend to adopt their partner's language variety more often than individuals from high-status social groups (Palomares, Giles, Soliz, & Gallois, 2016), and speakers tend to converge to partners from other social groups based on their positive or negative attitude towards that group (Babel, 2010).

However, this evidence comes mostly from naturalistic studies, where exercising experimental control of potential confounding variables is not a viable option. Thus, it is unclear whether these findings actually reflect social affiliation components in language accommodation, or if they are confounded by other variables, such as differences in interlocutors' actual language use. To understand the extent to which lexical entrainment implies a social affiliation component, it is therefore important to examine the extent to which increased lexical entrainment causally follows from individuals' increased social disposition towards social affiliation (e.g., after experiencing ostracism). And it is important to examine this causal relationship in an experimental setting where the partner's language use remains the same across experimental conditions.

In sum, a social affiliation account of lexical entrainment explains the phenomenon as an affiliation behaviour, i.e., as a way to increase social affiliation

with a partner. Evidence for this account comes mostly from group-level comparison studies showing that a speaker's mimicry of their partner's utterances increases their partner's affiliation behaviours, and from studies suggesting that speakers tend to accommodate their language more often in situations where they are generally expected to exhibit prosocial behaviours, such as when interacting with a person from a higher-status social group. However, the causal relationship between increased affiliation dispositions and lexical entrainment has yet to be tested.

## 2.3. What could lexical entrainment tell us about the variability of language use?

Lexical entrainment can be explained in terms of lexical processing that *is not* mediated by beliefs and in terms of lexical processing that *is* mediated by beliefs. Unmediated accounts suggest that the tendency for a speaker to entrain is mainly driven by lexical priming, in that the recent processing of a word enhances lexical retrieval and reuse. Mediated accounts, in contrast, suggest that the tendency to entrain is mediated by audience design and social goals. In particular, audience design accounts suggest that people entrain to a partner to enhance mutual comprehension, while social affiliation accounts suggest that people entrain as a way to increase their likeability and decrease social distance.

There is supporting evidence for each of these accounts. It is uncontroversial that speakers tend to reuse the same words as their partner, and we know that the likelihood of entrainment increases as the lag between prime and target decreases, pointing out the role of ease of lexical retrieval due to recent processing (i.e., priming effects) as an important lexical entrainment component (e.g., Branigan et al., 2011; Hopkins, Yuill, & Branigan, 2017). Moreover, the fact that speakers entrain more often to a partner who is presented as less capable versus a partner presented as more capable (Branigan et al., 2011) demonstrates that this behaviour is sensitive to audience design, at least under certain circumstances. In addition, evidence that a speaker's tendency to repeat a partner's utterances increases that partner's social affiliation behaviours (van Baaren et al., 2013) and the fact that speakers tend to

accommodate their language use in situations where they are expected to have increased social affiliation goals (Giles, 1973), suggest that lexical entrainment might also imply a social affiliation component; but the causal relationship between social affiliation and lexical entrainment has yet to be tested.

Thus, accounts highlighting lexical retrieval and accounts featuring beliefs do not need to be mutually exclusive, and it is in turn important to understand whether and how lexical processing interacts with speakers' beliefs and social affiliation goals during entrainment. This consideration positions lexical entrainment as an excellent test-case to explore the interplay between lexical processing and beliefs, and to understand how that relationship varies across not only social situations but also across individuals.

However, one of the first things that catches attention when reviewing lexical entrainment research is that most of lexical entrainment data comes from group-comparison studies, remaining uncertain whether (and if so, why) the degree of lexical entrainment varies from person to person. This is a critical question to understand how potential lexical entrainment mechanisms interact with each other, and by extension how these mechanisms influence the variability of lexical choices. In particular, lexical entrainment could vary across individuals depending on their lexical retrieval, audience design, and social affiliation skills. But the degree to which an individual entrains to a partner can also vary depending on, for example, how lexical retrieval interacts with audience design skills.

But to understand individual differences, it is first necessary to acknowledge and address the methodological challenges of individual differences studies. Given a long-lasting tradition of group comparisons in our field, most current psycholinguistic tests have been designed to distinguish between groups of participants, thus leading to the minimisation of differences between individuals. An important pre-requisite to study individual differences is  using tasks with high test-retest reliability (see Hedge, Powell, & Sumner, 2018), which captures the degree to which a task consistently elicits the same score for each individual across time: If we can show that a lexical entrainment task elicits the same within-individuals scores across time, then we can claim that the phenomenon reflects stable differences in

language processing across individuals. Only after doing that should we start exploring the extent to which lexical entrainment mechanisms underlie such potential differences.

Moreover, an important consideration about current theories of lexical entrainment is that it has not yet been tested whether the phenomenon actually follows from increased social affiliation. This is important to understand whether and how the variability of lexical choices is influenced by social affiliation, a question that has been somewhat ignored by the psycholinguistics of dialogue. To understand this relationship, it is necessary to test the causal relationship between increased social affiliation and speakers' tendency to lexically entrain. One way to manipulate individuals' increased social affiliation dispositions is by inflicting feelings of ostracism, a manipulation that has been already demonstrated to increase affiliation behavioural mimicry (Lakin et al., 2008).

Critically, understanding the effects of ostracism on lexical entrainment could also have important implications for theories of behavioural mimicry and language processing more generally, since it remains uncertain whether non-linguistic behavioural mimicry and linguistic mimicry (entrainment) are supported by the same mechanisms. Moreover, understanding the effects of ostracism on lexical entrainment can also help define the scope of ostracism effects on human behaviour; so far, ostracism has been thought to increase non-functional affiliation behaviours, such as foot-shaking (Lakin et al., 2008), but it remains unexplored if experiencing ostracism also influences functional behaviours, such as language use during social interaction.

Furthermore, most audience design research has centred around how language processing is affected by the interplay between speakers' language use and the information they share with a particular partner, developing numerous accounts of how personal common ground is established and used to achieve mutual understanding. However, we know very little about community-level influences on how speakers use language during dialogue (e.g., with a partner from their own community or from another community); in fact, we do not know much about how speakers establish community-level knowledge in the first place.

Language users are not just individuals acting in isolation – they form part of larger speech communities with shared patterns of language use. As such, it is important to understand how community-level experience affects individuals' language use during single linguistic encounters, and how single linguistic encounters help establish speakers' representations of community-level knowledge. In particular, communal common ground studies could reveal how individuals' previous linguistic experience may affect their language use during dialogue (e.g., do speakers entrain to the same extent to a partner from their community versus another community). Moreover, they could also cast light on how personal common ground with a particular partner updates community-level knowledge (e.g., do speakers generalise their lexical choices across two partners based on their speech communities?).

## 2.4. An empirical investigation of variation in language use through the lens of lexical entrainment

In four studies, this thesis aims to inform sources of variability in language use, by empirically investigating individual differences and situational factors in lexical entrainment. We first aimed at understanding the extent to which lexical entrainment can reliably capture individual differences in language processing, by examining the test-retest reliability of a web-based lexical entrainment task (based on Branigan et al., 2011) in two online studies (see Chapter 3). Then, we investigated which mechanisms may underlie individual differences in lexical entrainment, by looking at whether and how speakers' tendency to reuse their interlocutor's words is predicted by individual differences in schizotypy and age, factors that are thought to correlate with differences in lexical retrieval, audience design, and social affiliation (see Chapter 4).

Moreover, we aimed at investigating the causal relationship between social affiliation and lexical entrainment, by testing the effects of experiencing ostracism on speakers' tendency to entrain to a partner (Chapter 5). Importantly, we examined whether post-ostracism lexical entrainment effects were tailored to repairing a

particular social relationship with a partner or increasing social affiliation more generally, and whether such effects were mediated by individual differences in personality.

Furthermore, we aimed to understand community-level influences on how speakers vary their referential expressions and on how they establish community-level knowledge from single linguistic encounters (see Chapter 6). Critically, we moved beyond looking only at how beliefs could affect lexical entrainment with a particular partner, to instead use lexical entrainment as a way to examine how single linguistic experiences can influence speakers' language with subsequent partners. In particular, we first examined how speakers' tendency to entrain to a partner varied depending on that partner's speech community, and then examined how speakers' extrapolation of their own referential choices to a subsequent partner was modulated by the two partners' speech communities.

Finally, we discuss the implications of our findings for theories of lexical entrainment and the variability of language use, and outline future research directions to keep furthering our understanding of how language use works.

## Chapter 3

# 3. Lexical entrainment reflects a stable individual trait

The following chapter presents a scientific paper, titled *Lexical entrainment reflects a stable individual trait: Implications for individual differences in language processing,* published in the Journal of Experiment Psychology: Language, Memory, and Cognition, and written in collaboration with Dr Hugh Rabagliati and Prof Holly P. Branigan. The manuscript attached here is the accepted version of the article, and it includes its original abstract and list of references. Numbers of tables, figures, and sections were adapted to match the rest of this thesis.

As detailed in Chapter 1, the aim of this article was to understand the test-retest reliability of a lexical entrainment instrument that we used in all the studies of this thesis, in order to evaluate its ability to reflect individual differences in language processing.

# Lexical entrainment reflects a stable individual trait: Implications for individual differences in language processing

Anita Tobar-Henríquez, Hugh Rabagliati, Holly P. Branigan

University of Edinburgh

Author Note

This research was partly funded by CONICYT PFCHA/DOCTORADO BECAS CHILE/2015 - 72160212.

All analyses, data, and stimuli are available at https://github.com/anitatobar/Test-retest-reliability-of-lexical-entrainment-task.

Correspondence concerning this article should be addressed to Anita Tobar-Henríquez, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ.

Contact: anita.tobar@ed.ac.uk

## 3.0. Abstract

Language use is intrinsically variable, such that the words we use vary widely across speakers and communicative situations. For instance, we can call the same entity *refrigerator* or *fridge*. However, attempts to understand individual differences in how we process language have made surprisingly little progress, perhaps because most psycholinguistic instruments are better-suited to experimental comparisons than differential analyses. In particular, investigations of individual differences require instruments that have high test-retest reliability, such that they consistently distinguish between individuals across measurement sessions. Here, we established the reliability of an instrument measuring lexical entrainment, or the tendency to use a name that a partner has used before (e.g., using *refrigerator* after a partner used *refrigerator*), which is a key phenomenon for the psycholinguistics of dialogue. Online participants completed two sessions of a picture matching-and-naming task, using different pictures and different (scripted) partners in each session. Entrainment was measured as the proportion of trials on which participants followed their partner in using a low-frequency name, and we assessed reliability by comparing entrainment scores across sessions. The estimated reliability was substantial, both when sessions were separated by minutes and when sessions were a week apart. These results suggest that our instrument is well-suited for differential analyses, opening new avenues for understanding language variability.

Test-retest reliability, individual differences, lexical entrainment, language production, alignment

## 3.1. Introduction

Language use is intrinsically variable: People express themselves differently across different social situations (Eckert & Rickford, 2001; Gregory & Carroll, 2018), and across their lifespans (e.g., Juncos-Rabadán, Facal, Rodríguez, & Pereiro, 2010; March, Wales, & Pattison, 2011). Language use also shows wide variation from individual to individual, depending on demographic characteristics such as gender (e.g., Leaper & Ayres, 2007). This variability shows up even in references to commonplace objects, where speakers might variously refer to the same object as a *fridge, refrigerator* or even just *thing*. A range of theories have drawn on psycholinguistic experiments using group-level approaches to elucidate the cognitive and social factors that might inform and constrain how we make lexical choices, and in doing so they have also cast light on factors that can affect how we use language more generally. In this paper, we take a novel approach to investigating variability in language use, seeking to establish whether the way in which we make lexical choices reflects stable individual differences in how we process language. We focus on one particular determinant of lexical choices: **lexical entrainment**[2]**,** or speakers' overt use of names that their conversational partner has used before (e.g., to use *refrigerator* after hearing a partner use *refrigerator*; Garrod & Anderson, 1987; Brennan & Clark, 1996; Branigan, Pickering, Pearson, McLean, & Brown, 2011).

Research on lexical entrainment has revealed that lexical choices are affected by cognitive factors, such as memory, as well as social factors, such as the identity of a partner. However, there are still competing theories as to what underlies the phenomenon, and as to how cognitive and social factors interact in language

[2] This phenomenon is also sometimes referred as lexical alignment, to imply that the phenomenon of lexical entrainment is driven by two interlocutors' simultaneous activation of the same lexical representation, i.e., corresponding to a shared conceptualisation of a specific entity (Clark & Wilkes-Gibbs, 1986; Pickering and Garrod, 2004). We use the term lexical entrainment as we focus on speakers' overt reuse of their partner's lexical choices, and thus we do not consider the underlying mechanisms (e.g., lexical priming, Pickering & Garrod, 2004; audience design, Clark, 1996) that might give rise to this observable behavior.

production. For instance, Pickering & Garrod (2004) suggested that we reuse our partner's words mainly because we have recently processed them; recent processing makes the words more accessible in memory, and therefore facilitates retrieval and reuse (e.g. Neely, 1976; Meyer, 1996). By contrast, audience design accounts suggest that our tendency to reuse our partner's lexical choice occurs because we tend to adopt their perspective during conversation in order to facilitate mutual comprehension (Clark, 1996). Evidence for this account comes from demonstrations that we reuse lexical choices in a partner-specific fashion (Brennan & Clark, 1996; Horton & Gerrig, 2002, 2005). Other theories have suggested that reusing our partner's lexical choices can also be considered a pro-social behavior, as it makes us more likeable to our interlocutors (Van Baaren, Holland, Steenaert, van Knippenberg, 2003). Consistent with this, sociolinguistic studies have shown that social hierarchies and community membership correlate with our tendency to reuse a partner's choices. For instance, members of low status groups are more likely to entrain than members from high status groups (Palomares, Giles, Soliz, & Gallois, 2016). Similarly, people are more likely to entrain to more prototypical community members than to non-prototypical community members (Gallois & Callan, 1991).

In this paper, we make the case that studying individual differences in lexical entrainment can be informative not only about this specific phenomenon, but also about theories of language use more generally. For instance, given that lexical entrainment could be explained as a consequence of audience design, an individual differences' approach could be used to test the hypothesis that an individual's perspective-taking skills might predict the degree to which they engage in lexical entrainment (cf. Hopkins, Yuill, & Branigan, 2017), which would in turn cast light on the effects of perspective-taking in language processing more generally. Similarly, individual variability in lexical entrainment can also cast light on how situational factors might interact with individual traits factors, such as personality, in language use. For example, are individuals who are more prone to exhibit pro-social behavior (agreeableness) more likely to lexically entrain to a non-prototypical community member than individuals who are less agreeable? Finally, individual differences in how people entrain at the lexical level can inform theories of our general tendency to

reuse a partner's linguistic choice at other levels of structure (e.g., phonetics, syntax, etc.; Pickering & Branigan, 1998; Pickering & Garrod, 2004). If entrainment at different levels of linguistic structure is supported by a domain-general mechanism for imitation, then our tendency to reuse a partner's lexical choice should correlate with, for example, our tendency to reuse a partner's syntactic choice (e.g., passive versus active structures; see also Horton, 2014, for related research on individual differences in syntactic entrainment).

However, a pre-requisite for studying individual differences in phenomena such as lexical entrainment is possessing instruments and protocols that are calibrated to allow us to reliably measure these behaviors at the individual level. One important aspect of this is ensuring that instruments have high test-retest reliability, such that they can consistently distinguish between individuals across two measurement sessions (Shrout & Fleiss, 1979; Polit, 2015; Berchtold, 2016). Test-retest reliability is usually quantified by measuring the correlation coefficient between two sets of measurements from the same group of individuals; classically, the reliability of a test is said to be excellent if that coefficient is above .8, substantial if between .8 and .6, moderate if between .6 and .4, and poor if below .4 (Cicchetti & Sparrow, 1981; Landis & Koch, 1977). Understanding the test-retest reliability of an instrument is critical because it is a key determinant of the statistical power of a study. If studies are conducted using instruments that have a low test-retest reliability, then their ability to detect relationships with other constructs will be compromised by their inability to consistently distinguish between individuals on the dimension being measured. For instance, in order to investigate the relationship between agreeableness and lexical entrainment, we would need instruments that consistently distinguish between individuals in terms of both their degree of agreeableness and their propensity towards lexical entrainment.

Reliable instruments exist for testing certain individual differences, such as personality traits (e.g., Big Five; John & Srivastava, 1999). But recent work has suggested that many of the most well-known paradigms for assessing cognitive processing actually have poor test-retest reliability. Hedge, Powell, & Sumner (2018) demonstrated that a range of classic tasks, such as the Flanker and Stroop tasks,

which reliably elicit effects at the group level, do not reliably measure individual variation, producing test-retest reliability scores that often fail to reach even a moderate level. The reason for these low scores, Hedge et al. argue, is that these instruments produce a distinctive restricted range of responses, which minimise variability between respondents (e.g., almost all participants show a Stroop cost, and this cost is similarly-sized across participants). Although this feature is highly desirable for experimental research, it is problematic for correlational studies, which need to elicit a large enough range of scores to capture individual variation. The fact that reliable experimental tasks elicit minimally different effect sizes between individuals compromises the instruments' ability to distinguish between these individuals, and thus leads to low test-retest reliability.

This claim has important potential consequences for the study of language processing, because it seems quite likely that many experimental tasks that have been used for studying individual variation in this field are actually ill-suited for that purpose (Kidd, Donnelly, & Christiansen, 2018). For example, a number of studies have assessed individual variation in statistical learning, i.e., the ability to learn statistical co-occurrences of features in our environment (Siegelman, Bogaerts, & Frost, 2017). However, most of the tasks that have been used to assess individual differences in statistical learning were actually designed for studying group-level comparisons, and so it is not clear that they reliably measure individual variation. Indeed, there is some evidence that test-retest reliability in these tasks is compromised (Arnon, 2019; Pardo, Urmanche, Wilman, & Wiener, 2017; Siegelman & Frost, 2015). This point is important because, in principle, it casts doubt on whether prior findings are likely to replicate, as correlations drawn from instruments with low test-retest reliability are likely to be either false positives or negatives. Thus, it implies that we should be wary about using those studies to draw theoretical conclusions.

These considerations highlight the importance of establishing the test-retest reliability of an instrument before using it to study individual differences. Here we aim to establish the test-retest reliability of an instrument for measuring the phenomenon of lexical entrainment, both as a necessary step in the development of

sound correlational studies of the phenomenon, and also as a way to evaluate previous work that has examined individual variation in lexical entrainment. For example, Hopkins et al. (2017) studied lexical entrainment in a sample of typically developing and autistic children, and found that individuals' tendency to entrain did not correlate with measures of theory of mind or inhibitory control. However, without knowing the test-retest reliability of the lexical entrainment instrument, it is hard to interpret these null results. If the instrument has low reliability, then we should not typically expect to find reliable correlations between measures of entrainment and inhibitory control, even if the underlying factors are indeed associated. By contrast, if test-retest reliability is high, then these null findings are more likely to be indicative of true null associations. Thus, understanding the test-retest reliability of lexical entrainment instruments presents an important goal.

Branigan et al. developed a lexical entrainment instrument, which we will adapt in this paper, that has been repeatedly shown to elicit reliable effects at group-level, experimental comparisons (e.g., Branigan et al., 2011, 2016; Hopkins et al., 2017). In this instrument, participants collaborate with a confederate to match and name pictures. The experimental targets are pictures of objects that can be named with both a disfavoured and a favoured name (e.g., *brolly* versus *umbrella*, in British English); these materials have been pre-tested to ensure that participants rarely use the disfavoured name spontaneously, but still consider it an acceptable name for the object. In the main matching-and-naming instrument, participants always name the experimental targets after the confederate. Lexical entrainment is then measured as the proportion of trials on which participants use the same name used by the confederate. Importantly, Branigan and colleagues have consistently shown that individuals are more likely to use a disfavoured name (e.g., *brolly*) after the partner has used the disfavoured name (e.g. *brolly*) than after the partner has used the favoured name (e.g., *umbrella*) or compared to its baseline frequency of use (Branigan et al., 2011, 2016; Hopkins et al., 2017), and so they have demonstrated that this instrument elicits experimentally reliable entrainment effects for disfavoured names. Moreover, they showed that speakers' propensity to entrain to a partner's use of a disfavoured name was not affected by modality: Individuals were equally likely

to use a partner's disfavoured name during a written computerised interactive picture-naming task as during a spoken computerised version of the same task, providing evidence that effects elicited by this instrument generalise to speech (Branigan et al., 2011). But is the instrument also reliable for correlational studies?

Interestingly, this lexical entrainment instrument contains features that could enhance test-retest reliability. Its most critical feature is that it is designed to measure a general tendency to lexically entrain rather than to measure entrainment to a specific lexical item. To wit, each trial offers participants the opportunity to entrain to a different lexical item, which - we assume - means that behavior on that trial is relatively independent of behavior on previous trials (e.g., entraining to call a fridge a *refrigerator* should not in principle influence whether you call an umbrella a *brolly*). By contrast, typical paradigms used for measuring other types of entrainment contain design elements that might reduce test-retest reliability, because they are typically designed to measure entrainment to one feature only. For example, paradigms measuring the reuse of a partner's syntactic choice (syntactic entrainment) tend to assess participants' tendency to entrain to a specific syntactic structure (e.g., passive structures) and therefore necessitate that participants process the same linguistic structure repeatedly (e.g., Kaschak et al., 2011; Branigan & Messenger, 2016).

Measurements of entrainment to only one linguistic structure are likely to be quite strongly affected by participants' idiosyncratic experience with, or preference for, that specific structure, and thus may not be indicative of a general structure-independent tendency to syntactically entrain. Moreover, if a task repeatedly tests entrainment to a single structure, this would likely increase measurement error and, in principle, could lead to participants showing maximally large effects in the manner described by Hedge et al. (2017), which would leave little room for measuring individual differences. In principle, this issue could be surmounted if syntactic entrainment instruments used different syntactic structures in each critical trial, but such an approach would pose significant practical challenges. By contrast, it is simple to use different lexical items in each trial, such that lexical entrainment instruments measure entrainment anew in each trial, and thus do not fall prey to the criticisms of Hedge and colleagues.

### 3.1.1. The present studies

To investigate the stability of a lexical entrainment measure, we conducted two internet-based studies in which native speakers of British English engaged in two sessions of an interactive online picture matching-and-naming task. Our task was based on the task used in Branigan et al. (2011), and our materials were normed with a new, representative internet-based sample. In the main task, participants alternated turns with what they believed to be an online partner to either match or name a picture (in reality the 'partner' was always pre-programmed software). Given previous evidence that this task reliably elicits entrainment effects for disfavoured labels when participants have experienced the partner previously using a disfavoured label (Branigan et al., 2011, 2016; Hopkins et al., 2017), we measured entrainment to the use of disfavoured names only: Experimental trials comprised a target that, in British English, could be named with both a highly favoured name, e.g. *umbrella*, and a disfavoured, but acceptable, name, e.g., *brolly*, and the partner always used the disfavoured name to refer to the targets. Importantly, participants always matched experimental targets (i.e., responded to their partners naming the targets) before themselves naming the targets on a subsequent trial, and we measured entrainment as the proportion of trials on which the participant used the same disfavoured name as they had previously experienced the partner using.

In each study, we sought to establish first whether participants lexically entrained with an unseen partner in an online interactive picture naming-and-matching task, and second whether their propensity to lexically entrain was consistent across time. In our first study, we measured the test-retest reliability of lexical entrainment over a short time period: Participants completed two sessions immediately consecutively. Importantly, in each session, entrainment was measured with different items (e.g., *refrigerator/fridge* would be tested only in Session 1, while *brolly/umbrella* would be tested only in Session 2), meaning that the test-retest reliability should reflect an individuals' general tendency to lexically entrain, rather than their tendency to use particular (low-frequency) terms. In the second study, we

measured reliability over a longer time period, with sessions separated by 7-to-8 days.

In both studies, we assessed whether there was a group effect of lexical entrainment by comparing whether the disfavoured name was used more often in the main task than in a spontaneous picture naming-and-matching task that had been used to norm the materials (and that did not offer opportunities for entrainment). We then measured the test-retest reliability of our lexical entrainment measure in two ways. First, we calculated the relative rankings correlation between participants' use of disfavoured names in the first session and in the second session, assessing whether participants' degree of entrainment was ranked the same across sessions. Second, we measured the absolute consistency between participants' tendency to use disfavoured names in each session, in other words whether the instrument elicited exactly the same result for each participant in each session. In our design, we aimed to minimize situation-specific effects on lexical entrainment by using different stimuli across the two sessions, and by telling participants that they would be playing against different 'partners' in each of the two testing sessions, to avoid any possible partner-specific influences on lexical entrainment. Importantly, since individuals can have encountered disfavoured names in different proportions in previous experience, we aimed to minimize possible effects of past experience of the disfavoured names by using a range of 28 items, so that individual differences in previous experience with particular names could not explain participants' overall tendency to use disfavoured names during this task.

## 3.2. Materials' creation: Norming tasks

We conducted two norming tasks to create our experimental items, which comprised a target picture of an object that could be named with a favoured name in British English (e.g., *umbrella*) and a disfavoured, but acceptable name (e.g., *brolly*). Ethical approval for this norming procedure was obtained from the Psychology Research Ethics Committee of the University of Edinburgh (72-1617/9). In order to

create the pairs of favoured and disfavoured names for each experimental target, we conducted an initial pre-test with a different set of participants, drawn from the same population as those in the main studies. 60 native speakers of British English (aged 18-60, M=36, SD=11) answered two questions in an online survey (via Prolific). For each of 120 pictured objects, participants provided a favoured name for the picture (i.e., spontaneous naming, *What is the first word you would use to name this object?*), followed by a less-favoured name (i.e., forced naming, *What other word could you use to name this object?*).

From these ratings, we gathered 50 potential target pictures, for which at least 70% of participants had provided the same favoured name, and at least 15% of participants had provided the same disfavoured name. Importantly, the disfavoured names did not consistently come from specific registers or dialects of British English. The 50 potential targets were then entered into a second rating task, in which 60 new native speakers of British English (aged 18-60, M=38, SD=10) rated the acceptability of these disfavoured names with respect to the pictures on a scale from 1 to 7, where 1 corresponded to ‘Not acceptable at all’ and 7 corresponded to ‘Highly acceptable’. We used this to create the final set of 28 disfavoured names, each of which had an acceptability rating above 5.3 (M=6.1, SD=.5), and had been used with a frequency below 30% (M=7%, SD=7%).

We then split these items into two sets of 14 (see Tables 1 and 2) that were matched in acceptability (Set 1: M=6.2, SD=.5; Set 2: M=6.1, SD=.4) and frequency of use during spontaneous naming (Set 1: M=7.1%, SD=7.2%; Set 2: M=7.6%, SD=7.1%). Across participants, we counterbalanced which set was presented in the first session, and which in the second session. We also used the first rating task to choose 14 filler pictures, in which at least 80% of participants agreed on the same favoured name.

*Table 3.1. Item Set 1.*

| Disfavoured name (and favoured name) | Spontaneous naming (%) | Forced naming (%) | Acceptability score (1-7) |
|---|---|---|---|
| pillow (cushion) | 12 | 72 | 5.3 |
| musical instrument (accordion) | 0 | 18 | 5.5 |
| picture (painting) | 22 | 43 | 5.6 |
| make-up (lipstick) | 0 | 33 | 5.6 |
| silverware (cutlery) | 3 | 10 | 5.7 |
| flower (rose) | 1 | 91 | 6.0 |
| rodent (mouse) | 0 | 65 | 6.3 |
| loo (toilet) | 5 | 63 | 6.4 |
| mobile (phone) | 13 | 48 | 6.4 |
| refrigerator (fridge) | 2 | 48 | 6.5 |
| toad (frog) | 8 | 60 | 6.7 |
| aeroplane (plane) | 20 | 45 | 6.8 |
| memory stick (usb) | 11 | 29 | 6.8 |
| bicycle (bike) | 10 | 67 | 6.8 |

*Table 3.2. Item Set 2.*

| Disfavoured name (and favoured name) | Spontaneous naming (%) | Forced naming (%) | Acceptability score (1-7) |
|---|---|---|---|
| biro (pen) | 3 | 34 | 5.3 |
| computer (laptop) | 8 | 83 | 5.6 |
| rowboat (boat) | 3 | 23 | 5.7 |
| fag (cigarette) | 10 | 53 | 6.0 |
| spectacles (glasses) | 4 | 45 | 6.0 |
| coach (bus) | 0 | 30 | 6.0 |
| nectarine (peach) | 5 | 15 | 6.1 |
| hat (cup) | 28 | 52 | 6.1 |
| hen (chicken) | 5 | 43 | 6.2 |
| brolly (umbrella) | 12 | 45 | 6.3 |
| bunny (rabbit) | 15 | 63 | 6.4 |

| | | | |
|---|---|---|---|
| pistol (gun) | 0 | 48 | 6.4 |
| inflatable ball (ball) | 1 | 26 | 6.5 |
| bathtub (tub) | 3 | 25 | 6.7 |

## 3.3. Study 1: Short-term reliability

Study 1 investigated whether individual levels of lexical entrainment could be reliably measured in two sessions a few minutes apart. Participants completed a picture matching-and-naming-task. On each trial participants were shown two images and, while alternating turns with an alleged partner, they either named or selected one of the pictures. On critical trials, we measured whether participants reused a disfavoured name that their partner had used earlier in the study.

### 3.3.1. Method

Ethical approval for this study was obtained from the Psychology Research Ethics Committee of the University of Edinburgh (72-1617/9).

**Participants.** We recruited 60 participants online using the portal Prolific [https://prolific.ac/]. To be included, participants had to be native speakers of British English, born and raised in the United Kingdom, and aged 18-60 (M=36, SD=12). Participants were paid £2.

**Procedure.** Participants completed two sessions of a matching-and-naming-task, each of which contained 28 matching trials and 28 naming trials. On each trial, participants were shown two pictures (Figure 3.1), and they then either clicked on the target picture named by their partner (matching trials) or typed the name of the indicated target picture (naming trials). Half of the trials were filler trials, on which the target picture only had a single name (e.g., *onion*). The other half were experimental trials, on which the target picture could either be named with a highly-

favoured name (e.g., *umbrella*) or a less-favoured but still acceptable name (e.g., *brolly*). Thus, each session used 14 experimental items and 14 filler items, meaning that participants completed 28 experimental items and 28 filler items in total.

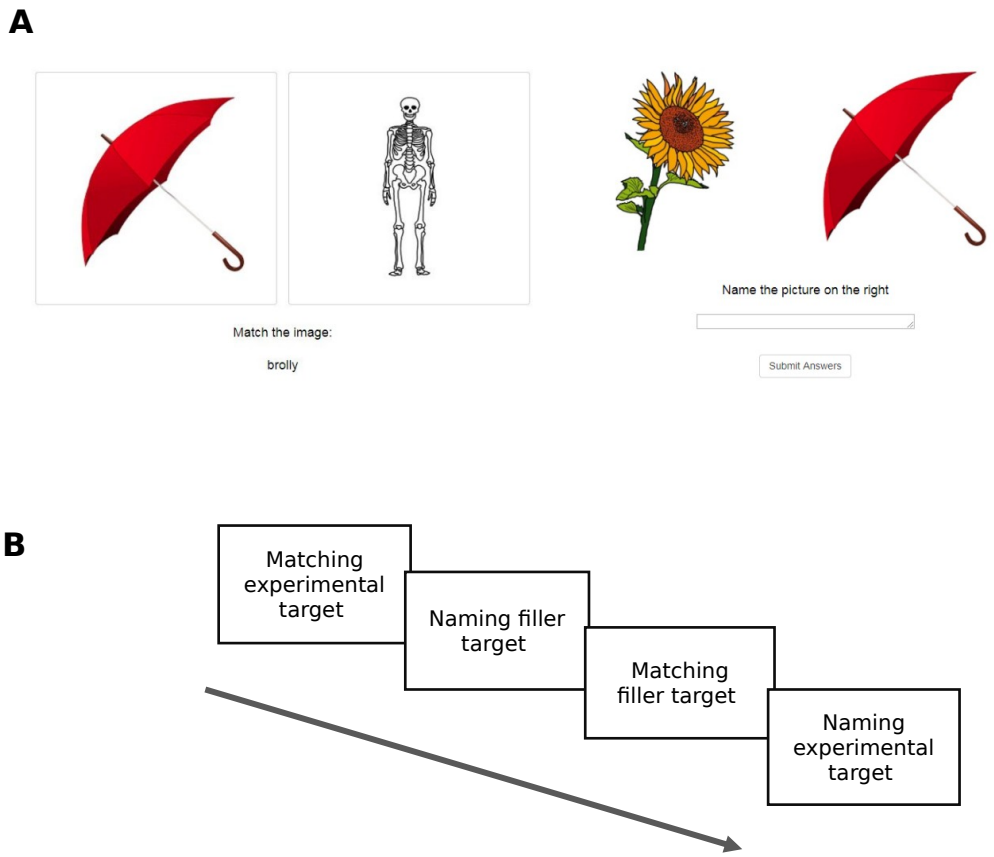


**Figure 3.1. A**. Examples of matching (left) and naming (right) trials (where the favoured word is *umbrella* and disfavoured is *brolly*). In matching trials, the participant selected the named target picture. In naming trials, they named the target. Targets were presented along with randomly selected distractors. **B. S**equence of experimental item and filler presentation. Participants first matched an experimental target with the corresponding disfavoured name, they subsequently named a filler, matched a filler, and finally named the previously matched experimental target.

The structure of the matching and naming task is illustrated in Figure 3.1. Participants alternated matching and naming trials with a ‘remote player’, who was in

fact pre-programmed software that provided scripted answers. The trial order was fixed and the latency between matching experimental target and naming experimental target was always 3 trials, as in Figure 3.1b. Importantly, the trial structure meant that the software 'partner' always named the experimental targets before the participants, using the disfavoured names exclusively (see Figure 3.1).

Participants were recruited to take part in this study on Prolific, using an advertisement that was visible only to individuals who met our inclusion criteria (see above). The advertisement stated that participants would play two sessions of a picture matching-and-naming task, and that they would play with a different remote player in each session. Prolific users interested in participating in the study were redirected from Prolific to a Qualtrics survey. After filling in an online consent form, they were told to wait to be matched with a remote player and, after two minutes, they were redirected to the first task (programmed with JSPsych and available at https://github.com/anitatobar/Test-retest-reliability-of-lexical-entrainment-task; de Leeuw, 2015), where they were asked to alternate turns with their partner to match and name one out of two pictures that would appear on the screen.

On each trial, participants saw two pictures and were asked to either wait for their partner's response so that they could select the correct (matching) picture, or to name the picture on the right/left (depending on where the target appeared, which was randomized) (see Figure 3.1). After matching and naming the 14 experimental items, they were told to wait to be matched to a new remote player. After two minutes, they were told the new partner was waiting for them and were asked to press a key to start the task. During the second session, participants matched and named 14 new experimental items. At the end of the task, participants were redirected to a second Qualtrics survey, where we checked participants' beliefs about the nature of their partner by asking *How many people did you play with during the two naming tasks?*; we coded whether participants reported playing with multiple partners, or explicitly indicated that they suspected they had played with a computer. Finally, participants were redirected to a Prolific website and received a completion code in order for us to confirm their payment.

### 3.3.2. Results

**Data processing and exclusions.** We coded all naming trials for whether they showed lexical entrainment (repeating the disfavoured name used by the partner) or not (using any other name). Occasionally, participants named or selected the distractor instead of the target; these trials were coded as NA. We excluded five participants because they reported believing that they had not played with a real person.

**Analyses.** We conducted two analyses, using the open source R language and environment (R Core Team, 2018) in RStudio (RStudio Team, 2015). All analyses and data can be found at https://github.com/anitatobar/Test-retest-reliability-of-lexical-entrainment-task.

We began by testing for the presence of a lexical entrainment effect, by comparing the percentage of disfavoured names used in our matching-and-naming tasks to the percentage of disfavoured names used in our first norming task. To do so, we used a paired-samples Wilcoxon test over the percentage of use of disfavoured names in each task.

Next, we assessed the test-retest reliability of the task by comparing the proportion of trials on which participants entrained in each session. In all analyses, the variables of interest were logit transformed proportions: We used this transformation over binary proportions to approach normality. In our first analysis, we used a Pearson's correlation to assess the degree to which the instrument could replicate the same ordering between respondents in the two sets of measurements. In our second analysis, we used intra-class correlation coefficients (ICC) to assess whether the instrument could also elicit the same exact result for each individual on each session. The ICC reflects the consistency between two or more raters (in this case, measurement sessions) for the same set of participants (Shrout & Fleiss, 1979; Polit, 2015; Berchotold, 2016), and its values fall between 0 and 1, with an ICC of 1 reflecting perfect consistency.

We calculated ICC values adopting two different approaches. First, we used an ANOVA-based approach estimating components of variance (McGraw & Wong,

1996; Shrout & Fleiss, 1979). Following Koo & Li (2016), we used a single-rating, absolute-agreement, two-way random effects model with two raters (testing sessions) across 55 individuals. In the commonly cited Shrout and Fleiss (1979; see also McGraw &Wong, 1996) nomenclature, this corresponds to ICC (2,1), which is sensitive to differences between session means. We used a two-way model because the sets of experimental items were counterbalanced across testing sessions, making participants' scores from Session 1 and Session 2 interchangeable. Moreover, we used a single-rating ICC type to compare each participant's tendency to entrain in Session 1 against their tendency to entrain in Session 2, rather than comparing Session 1 and Session 2 scores as a whole. In addition, we used an absolute agreement ICC definition (instead of a consistency definition) because we were not only interested in measuring ranking consistency across time but, most of all, we wanted to test our measure's ability to provide identical results in each measurement session.

Second, we calculated the ICC using a generalised mixed-effects models approach, which allows the calculation of standard error via bootstrapping (Nakagawa & Schielzeth, 2010). In particular, we built a two-way random effects model using logit transformed proportions as the independent variable. The model included random intercepts for both participants and sessions, and we used 1000 bootstrapping iterations to calculate ICC values and 95% confidence intervals.

In interpreting our reliability results, we adhere to conventional standards for judging test-retest reliability in correlational research: excellent or clinically required (.8), good/substantial (.6), and moderate (.4) (Cicchetti & Sparrow, 1981; Landis & Koch, 1977; see Koo & Li (2016) for discussion). However, it is important to note that setting explicit standards for judging these values is difficult, since the appropriateness of a coefficient depends on factors such as the purpose for measuring the reliability of the instrument, the time interval between measurements, the types of sample being used, and whether the underlying phenomenon is believed to be volatile (Crocker & Algina, 1986).

**Lexical entrainment effect.** We found strong evidence for lexical entrainment. On average, participants used the disfavoured names on 36% of critical

naming trials (SD=28%) across the two sessions. The percentages of use of disfavoured names during the matching-and-naming task were significantly higher than the percentages of use of these names during the spontaneous naming task used to norm the materials (M=7%, SD=7%, V=1, p<.0001), suggesting the presence of an entrainment effect (see Figure 3.2).

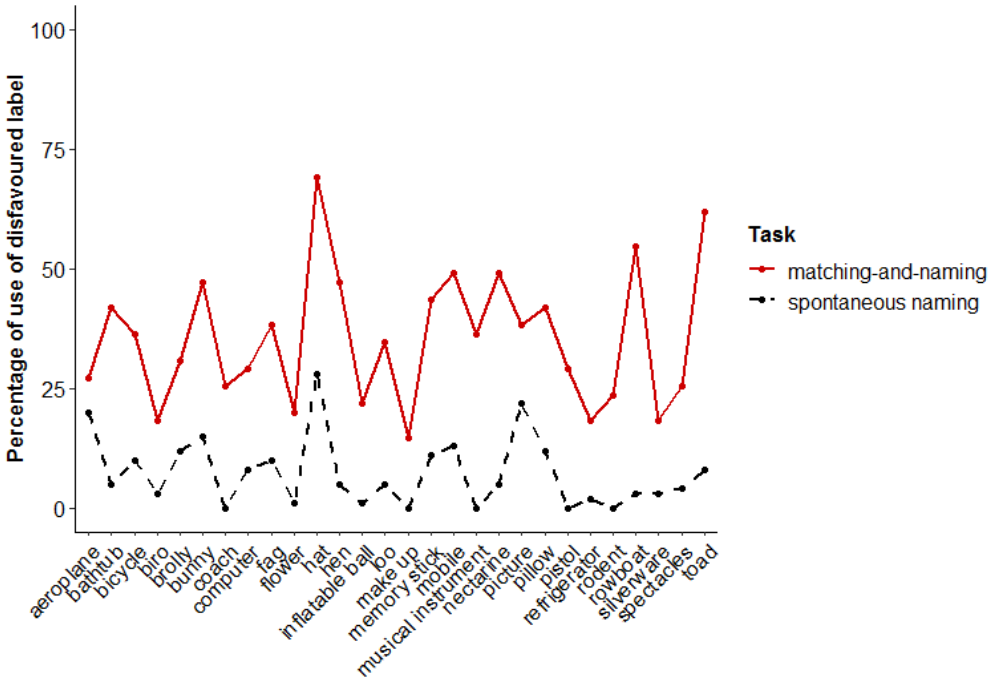


**Figure 3.2**. Percentage of use (y-axis) of disfavoured names (x-axis) in Study 1. The black (dashed) line represents the percentage of use of the disfavoured name in the spontaneous naming-task, while the red (solid) line represents the percentage of use of the disfavoured name in the primed matching- and naming-task used to measure lexical entrainment.

**Test-retest reliability.** Participants used disfavoured names on 34% of critical trials (SD=28%) in Session 1, and on 37% of critical trials (SD=28%) in Session 2. Importantly, our data shows a fairly substantial range of inter-individual variation in the degree of lexical entrainment (see Figure 3.3), which is an important prerequisite for correlational research. Moreover, we found a significant positive correlation between individuals' rates of lexical entrainment in Session 1 and lexical entrainment

in Session 2 (*r*=.73, p<.0001; 95% CI [.57, .83]). The ANOVA-based approach indicated an ICC value of .73 (p<.0001) with a 95% confidence interval between .57 and .82; this was confirmed by the generalised mixed effects models approach, which showed an ICC of .73 (p<.0001) with a 95% confidence interval between .57 and .83, and a standard error of .065. Importantly, a Levene Test revealed that our model did not show violation of the assumption of homoscedasticity across Session 1 and Session 2 (F=0.07, p=0.8), suggesting that inter-individual variation was similar across measurement sessions. Taken together, these results suggest that the two sets of measurements were not only correlated but also highly consistent at the individual level.

Thus, Study 1 shows that lexical entrainment is an effect that can be reliably elicited at the group level, across a range of items, even in this novel on-line task in which participants believed they were interacting with a remote partner. More importantly, these results also suggest that the test-retest reliability of this lexical entrainment task is quite substantial over a short time window, implying that lexical entrainment shows short-term stability within individuals, and that this task is well-suited for studying individual variation in language processing. Next, we assess the test-retest reliability of our task over a considerably longer time window.

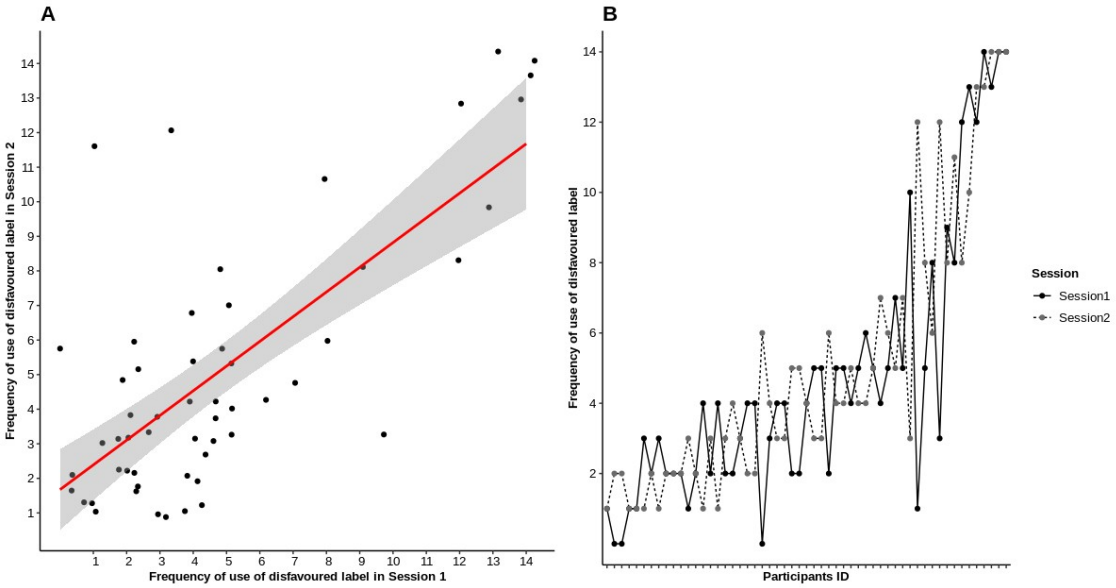


**Figure 3.3. A**. Positive correlation (*r*=.72) between the frequency of use of disfavoured names in Session 1 (x-axis) and Session 2 (y-axis). Points are jittered. The red line represents a linear regression between participants' scores in each session, while the grey shadow corresponds to a non-parametric regression smooth. **B**. Individual participants' scores in Session 1 (black, solid line) and Session 2 (grey, dashed line).

## 3.4. Study 2: Over-a-week reliability

Study 1 demonstrated substantial test-retest reliability of our lexical entrainment instrument when the testing sessions occurred with a gap of two minutes between sessions. In Study 2, we investigated whether test-retest reliability remains high when a seven-to-eight day gap is introduced between sessions.

### 3.4.1. Method

Except where detailed, Study 2 used the same methods as Study 1. Ethical approval for this study was obtained from the Psychology Research Ethics Committee of the University of Edinburgh (72-1617/9).

**Participants.** Study 2 used 60 further participants, aged 18-60 (M=31, SD=8), who were recruited using the same inclusion criteria as in Study 1.

**Materials and procedure.** Participants were recruited on Prolific, using an ad only visible to individuals who met our inclusion criteria (see above). The ad stated that participants would play two sessions of a picture matching-and-naming task, making explicit that the first session would take place immediately and the second session would take place in a week's time. The ad also stated that they would play with a different remote player in each session. Prolific users interested in taking part in Session 1 completed the same procedure as Study 1, except that they received a completion code to be paid for their participation in Session 1 immediately. A week later, they were served a new ad, allowing them to participate in Session 2, for which they followed the same procedure as Study 1. The second task was available to be answered for 48 hours. Those participants who wanted to participate in Session 2 were redirected to the second task and followed the same procedure as Study 1. Importantly, participants completed the manipulation check only once, at the end of Session 2, in order not to draw attention towards the nature of the remote partner before Session 2 was completed.

### 3.4.2. Results

**Data processing and exclusions.** Coding and exclusions were performed as in Study 1. We excluded ten participants because they did not complete the second session. Another five participants were excluded because they reported believing that they had not played with a real person.

**Lexical entrainment effect.** Again, we found strong evidence for an entrainment effect. On average, participants used the disfavoured names on 41% of trials (SD=24%) across the two sessions. The percentage of use of disfavoured names during the matching-and-naming task was significantly higher compared to the

spontaneous naming task (M=7%, SD=7%, V=1, p<.0001), suggesting the presence of an entrainment effect (see Figure 3.4).

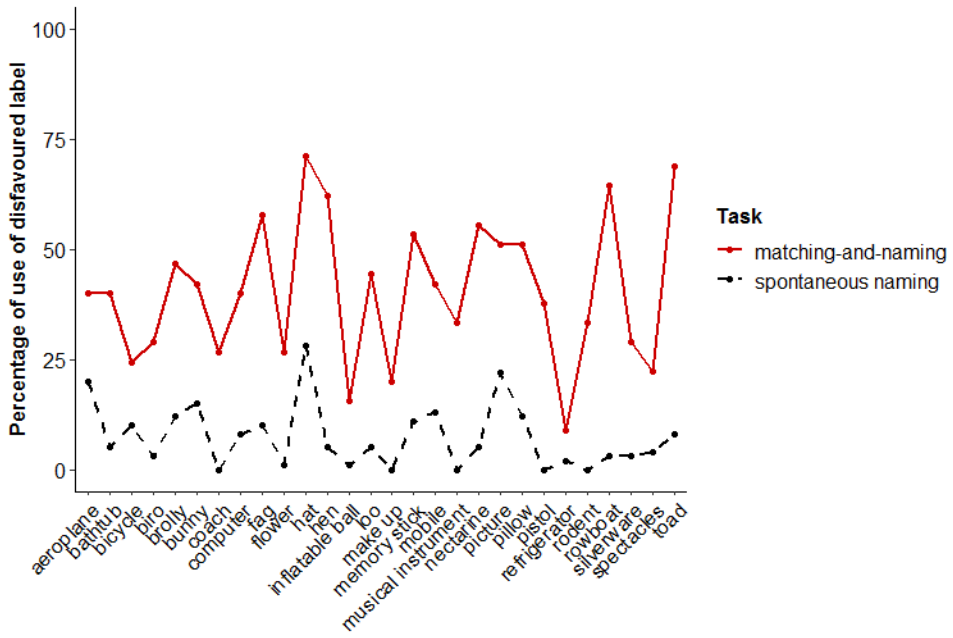


**Figure 3.4**. Percentage of use (y-axis) of disfavoured names (x-axis) in Study 2. The black (dashed) line represents the percentage of use of the disfavoured name in the spontaneous naming-task, while the red (solid) line represents the percentage of use of the disfavoured name in the primed matching- and naming-task used to measure lexical entrainment.

**Test-retest reliability.** Participants used disfavoured names on 40% of naming trials (SD=26%) in Session 1 and on 42% of naming trials (SD=26%) in Session 2. As in Study 1, our data shows a fairly wide range of inter-individual variation in the degree of lexical entrainment (see Figure 3.5). Moreover, we found a significant positive correlation between individuals' rates of lexical entrainment in Session 1 and lexical entrainment in Session 2 r=.61, p<.0001; 95% CI [.38, .77]). The ANOVA-based approach indicated an ICC value of .61 (p<.0001) with a 95% confidence interval between .4 and .77, and this was confirmed by the generalised mixed effects models approach, which also showed an ICC of .61 (p<.0001) with a

95% confidence interval between .41 and .76, and a standard error of .09. Importantly, the error terms of the generalised model were homogeneously distributed across Session 1 and Session 2 (F=0.17, p=0.7), suggesting that inter-individual variation was similar across measurement sessions. As in Study 1, these results suggest that the two sets of measurements were not only correlated but that they were also substantially consistent at the individual level. Although the confidence interval width was larger in Study 2 than in Study 1, it is important to note that Study 1's confidence interval is contained within Study 2, which suggests that lexical the test-retest reliability of our lexical entrainment task remains similarly high in the short- and the long-term.

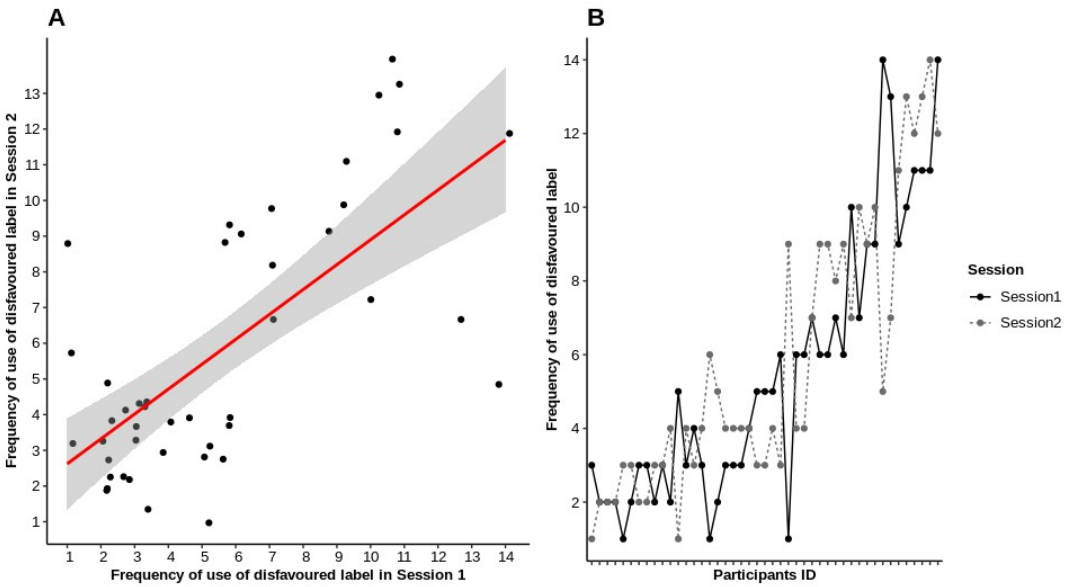


**Figure 3.5. A**. Positive correlation (*r*=.61) between the frequency of use of disfavoured names in Session 1 (x-axis) and Session 2 (y-axis). Points are jittered. The red line represents a linear regression between participants' scores in each session, while the grey shadow corresponds to a non-parametric regression smooth. **B**. Individual participants' scores in Session 1 (black, solid line) and Session 2 (grey, dashed line).

The results from Study 2 replicate the finding that lexical entrainment is an effect that can be reliably elicited at the group level, across a range of items, when participants believed they were interacting online with a remote partner. It also shows that our lexical entrainment task reaches a substantial level of test-retest reliability over a seven-to-eight day period. However, although the reliability of our task is still substantial over a week, it is lower than in the first study, which suggests that lexical entrainment may be influenced by situational factors.

## 3.5. Follow-up Analyses

The analyses reported above demonstrate that our lexical entrainment task elicits robust lexical entrainment effects at the group level, and reaches a substantial level of test-retest reliability not only across sessions separated by only two minutes, but also across sessions separated by seven-to-eight days. However, the ICC coefficient was slightly higher in the short-term (Study 1) than in the long-term (Study 2). Although Study 1's confidence interval is contained within Study 2, which suggests that the test-retest reliability of our lexical entrainment task is substantial in the short- and the long-term, in this section we report an additional analysis assessing the overall within-participants reliability of our task across results from both Study 1 and Study 2 together.

To do this, we used a Pearson correlation and both an ANOVA-based and a mixed effects model approach. To account for possible variances in the data explained by reliability differences between Study 1 and Study 2, we assessed the overall reliability of our task using an adjusted mixed-effects model, which included Study as fixed effect (sum contrast coded, i.e., -1/1). Across the two studies, participants used disfavoured names on 37% of critical trials (SD=27%) in Session 1, and on 39% of critical trials (SD=27%) in Session 2. We found a significant positive correlation between individuals' rates of lexical entrainment in Session 1 and in Session 2 ($r=.68$, $p<.0001$; 95% CI [0.56, 0.78]). The ANOVA-based approached indicated an obtained ICC value of .68 ($p<.0001$) with a 95% confidence interval between .56 and .78, and this was confirmed by a generalised mixed effects model

approach, which also showed an ICC value of .68 (SE=.06, p<.0001) with a 95% interval between .54 and .77. As expected, the error terms of the generalised model were homogeneously distributed across Session 1 and Session 2 (F=.01, p>.05), suggesting that inter-individual variation was similar across measurement sessions. Critically, the error terms of our model were also homogeneously distributed by Study (F=.14, p>.05), suggesting that inter-individual variation was similar in Study 1 and Study 2. These results, taken together with the finding that Study 1's CI is contained within Study 2's CI, support that our lexical entrainment task reaches a substantial level of test-retest reliability both in the short-term and in the long-term.

In further follow-up analyses, we addressed three critical features that could have affected the precision of measurement of our task: 1) sample size used in each study, 2) number of critical trials in our task, and 3) measurement independence between trials.

Although our samples are larger than many previous test-retest reliability studies in psychology and psycholinguistics (e.g., Larson, Baldwin, Good, & Fair, 2010; Strauss, Allen, Jorgensen, & Cramer, 2005; Arnon, 2019), it nevertheless is important to understand whether the precision of our estimates could be substantially increased if our sample size had been much larger, at least within the realms of practical possibility (cf. Schönbrodt & Perugini, 2013). For instance, if we had recruited twice the number of participants we used, would the precision of our measurement be substantially improved? Or, if we had used only half of the participants we recruited, how precise would our measurement be? To interrogate how sample size affected the precision of our estimates, we conducted a resampling analysis examining how test-retest reliability between Session 1 and Session 2 changed, as we varied the number of participants in each study. This thus provided us with a window into the effect of varying sample size, up to the limit of our instrument's measurement error.

To do this, we repeatedly compared the by-participant correlation across Session 1 and Session 2 over subsets of between 15 participants and the total number of participants in each study (Study 1 = 55, Study 2 = 45). We did this 1000 times for each sample size, drawing participants with replacement, and we used the resulting

data to calculate mean correlations across sessions, and to estimate 80% confidence intervals. The resulting Pearson’s correlations and their 80% confidence intervals are illustrated in Figure 3.6. By analysing and extrapolating from these data, we can assess the effects of sample size. As expected, increased sample size leads to narrower confidence intervals in both studies. However, the benefit of increasing the sample size seems to level off almost completely within the scope of our resampling analyses. Although this level off may result in part from the limits imposed by the measurement error of our task, they also suggest that increasing our sample size within the limits of what is practically possible for a psychology study would not change the overall interpretation of our retest reliability results. In particular, in Study 1, by about 35 participants the confidence interval's lower bound stabilises around .55 and the higher bound stabilises around .8, suggesting that with only 35 participants we would have found very similar results to the ones we found using 55 participants, which in turn suggests that increasing the number of participants would not make a difference in the interpretation of our short-term reliability results. In Study 2, by about 35 participants the confidence interval already ranges between around .4 and .77, suggesting that a sample size of 35 participants would provide a similar long-term reliability estimate to the one we obtained with an n of 45, which in turn suggests that increasing the number of participants would not affect the interpretation of our long-term reliability results either. Taken together, these results suggest that increasing our sample size would not have a substantial impact on the interpretation of our results: Test-retest reliability is substantially stable both in the short- and the long-term. However, our resampling analyses show that, as expected, a larger sample size would provide a more precise measurement of lexical entrainment, which would of course be helpful for future work.

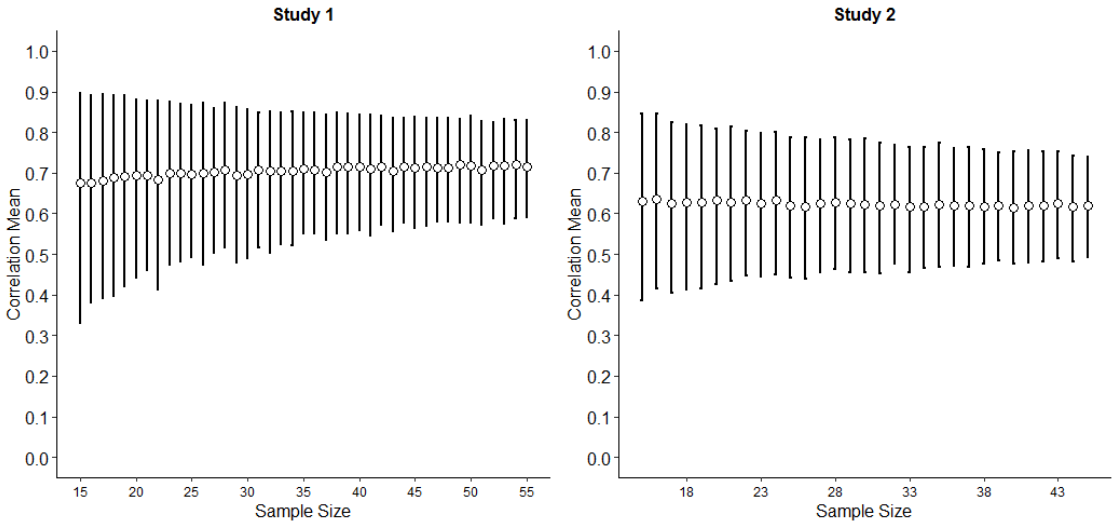


**Figure 3.6.** Mean correlations and 95% confidence intervals between participants' entrainment score in Session 1 and Session 2 (y-axis) for increasing number of sample size (x-axis).

Second, and relatedly, we addressed how many binary responses (or trials) are actually necessary to adequately capture the degree to which an individual tends to entrain. For example, if our fourteen-trial instrument were instead twice as long, would the benefit of enhanced measurement outweigh the cost of the additional trials? Or if our instrument were only half as long, how accurate would our measurements be? To provide initial answers to these questions, we conducted a second resampling analysis, now examining how test-retest reliability between Session 1 and Session 2 changed as we varied the number of trials in Session 1. This provided us with a window into the effect of varying trial number, up to the limit of Session 2's measurement error.

To do this, we repeatedly compared the by-participant correlation across Session 1 and Session 2, but where Session 1 scores were now calculated for randomly sampled subsets of between 1 and 14 trials. We did this 1000 times for each number of trials, drawing trials by participant without replacement, and we used the resulting data to calculate mean correlation across sessions, and to estimate 95%

confidence intervals. The resulting Pearson's correlations are illustrated in Figure 3.7, and show that, for both studies, reliability increased as the number of trials increased.

Importantly, Figure 3.7 suggests that different numbers of trials are needed to reach the same levels of short-term vs. long-term reliability. In particular, a substantial level of short-term reliability (a correlation greater than .6) can be reached with a relatively small number of trials: By about 8-to-10 trials, the confidence intervals in Study 1 no longer range less than .6, and thus a short 8-to-10 trial instrument may be appropriate for studies that require only somewhat precise measurement. However, Figure 3.7 also shows that by about the same 8-to-10 trials, long-term reliability reaches only a moderate level in Study 2, i.e., confidence intervals no longer range below .4. In addition, Figure 3.7 shows that the benefits of increasing the number of trials seem to decrease as the number of trials gets higher in each study, and they start to level off by about 10-to-11 trials in both studies, but there is no absolute level off within our sampling scheme. Although this suggests that the reliability of our instrument might in principle keep increasing if the number of trials reaches over 14, it also suggests that the benefits of adding more trials are not likely to change the interpretation of our results, i.e., that lexical entrainment is a substantially stable behavior both in the short-term and in the long-term, and that the long-term reliability of lexical entrainment is only slightly lower than its short-term reliability. However, these results highlight that although a short 8-10 trial instrument may be appropriate for studies that require only somewhat precise measurement of lexical entrainment, longer instruments might be better-suited for correlational studies in general, and individual differences in particular.

Moreover, these results suggest that an instrument with more trials would be useful to distinguish particular individuals who scored close to each other, providing a more precise measurement of their tendency to lexically entrain. For instance, inspection of Figures 3b and 5b suggests that a fairly large proportion of participants entrained only a small number of times, and thus an instrument with a greater number of trials could be useful to distinguish these particular individuals, providing a more precise measurement of their tendency to lexically entrain, both in the short- and the long-term.

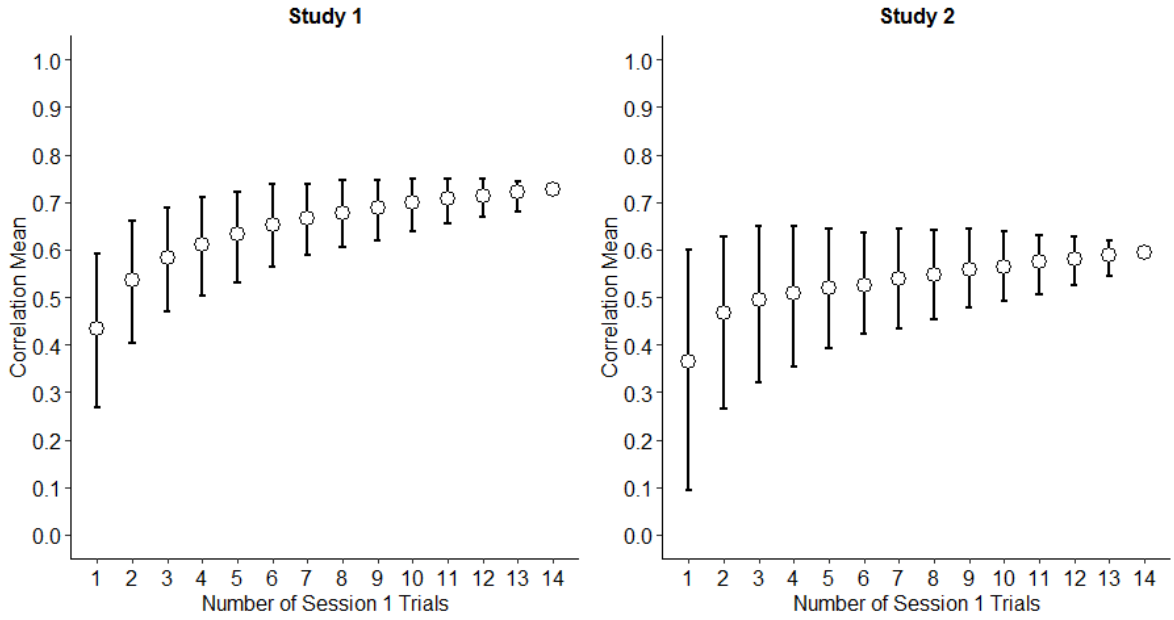


**Figure 3.7.** Mean correlations and 95% confidence intervals between participants' entrainment score in Session 2 and participants' entrainment by Session 1 trials (y-axis) for increasing number of Session 1 trials (x-axis).

Third, we additionally investigated the measurement independence between trials. We have argued that a critical methodological feature of this instrument is that each trial provides a relatively independent measure of participants' tendency to lexically entrain, such that whether participants entrained on one trial should not have a direct causal effect on the likelihood that they would entrain on the next trial. This is because we used different items on each trial (so that participants never entrained to the same name twice), and because the low-frequency names that we used were not drawn from any particular dialect or register (e.g., there is no British English dialect or register that standardly uses *brolly* for umbrella and *pillow* for cushion), so that there was no higher-order reason for using a low-frequency name.

To evaluate this argument we regressed entrainment against trial number for each study. Crucially, if trials are indeed independent of one another, then the degree of lexical entrainment should not vary strongly over the course of the experiment. Trial number was entered as a fixed effect (values were centred and standardized),

and participant and items were treated as random effects. We included by-participants random slopes for trial number. As Figure 3.8 shows, there was a slight numerical tendency for reduced entrainment across each Study, but the effect of trial number on participants' tendency to use disfavoured names was not significant in either study (Study 1: $\beta$=-0.27, p=.22, $\chi 2(1)$=1.47, p=.22; Study 2: $\beta$=-0.35, p=.055, $\chi 2(1)$=3.37, p=.06)[3]. These results are thus consistent with behavior on each trial being relatively independent of behavior on previous trials, and thus suggest that individual-level measurement error did not increase throughout the task as a function of trial dependence. This finding is thus consistent with the claim that our instrument succeeded in measuring individual-level behavior, because each trial provided a relatively independent instance of measurement, and in turn supports that our study design is well-suited for correlational research.

[3] Importantly, note that these results also rule out any possibility that participants were primed to use low-frequency words more generally by their 'partner's' use of a high proportion of low frequency words (i.e., the disfavoured names for experimental items). If participants were learning to use low frequency words over the course of the study, then we would expect them to show increasing entrainment as the study progressed. But instead, we found that entrainment rates gradually declined over the course of the study.

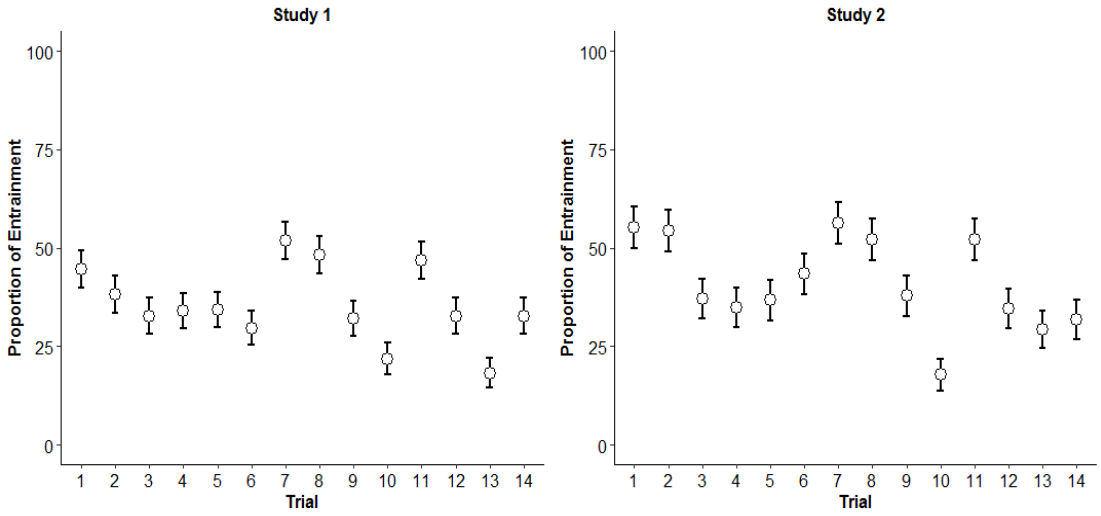


**Figure 3.8.** Percentage of lexical entrainment and standard error (y-axis) by trial number (x-axis). The figure includes trials from Session 1 and Session 2 for each study.

## 3.6. Discussion

Language use is variable not only within individuals but also between individuals. Studying individual variation in language processing is critical for understanding language use, but presents important methodological challenges that are quite distinct from the challenges posed when conducting psycholinguistic experiments. In particular, one important concern is that the test-retest reliability of some experimental psycholinguistic tasks may not be sufficient to allow the study of individual variation. It is therefore necessary to establish their level of consistency in advance.

In this paper, we have established the consistency of an instrument for studying lexical entrainment, which is the tendency to reuse a partner's lexical choice. Participants completed two sessions of a task in which they alternated turns with a partner to match and name pictures with two possible names. Participants always named the pictures after the partner, who used only disfavoured names, and we measured lexical entrainment as the proportion of trials on which participants

repeated the disfavoured names. In Study 1, the two measurement sessions took place two minutes apart; in Study 2, they took place seven-to-eight days apart. We found robust evidence for strong lexical entrainment in both studies. Although the disfavoured names had a frequency of use of 7% in a spontaneous naming-task, participants in our main experiment used disfavoured names 36% of the time in Study 1 and 41% of the time in Study 2. But more importantly, the studies reported here demonstrated that lexical entrainment was relatively consistent across sessions, both in terms of participants' relative ranking and in terms of their absolute tendency to entrain. In particular, individuals' tendency to lexically entrain reached a substantial level of reliability both in the short-term (i.e., across sessions separated by two minutes) and in the long-term (i.e., across sessions that were separated by seven-to-eight days). This pattern suggests that lexical entrainment has promise as an instrument for carrying out not only group-level experimental research but also individual-level correlational research.

What can these key findings – that lexical entrainment is robust, and that this effect is substantially stable within individuals – tell us about entrainment as a general phenomenon? Although our current results are not informative about exactly what mechanisms underlie lexical entrainment in this specific context (e.g., audience design, priming, pro-sociality), they provide evidence that these mechanisms contribute to making lexical entrainment a substantially stable behavior when the communicative situation is held roughly constant, as in our studies. Critically, the finding that the reliability of lexical entrainment reaches a substantial level even across 7-to-8 days suggests that inter-individual variability in the degree of lexical entrainment is underlain by stable individual differences, although what those traits are remain to be seen. For example, under the assumption that lexical entrainment interacts with social affect and pro-sociality (e.g., Van Baaren, 2003; Palomares et al., 2016 ), a person with a high propensity towards pro-sociality (agreeableness) may in general be more likely to lexically entrain than a person with a high propensity to feel very anxious in social situations (neuroticism) (e.g., see Gill, Harrison, & Oberlander, 2004, for individual differences in interpersonal syntactic priming). Alternately, assuming that lexical priming effects underlie lexical entrainment (e.g.,

Pickering & Garrod, 2004), individuals who are more susceptible to lexical priming may exhibit a greater basal tendency to reuse a partner's lexical label (i.e., a recently processed lexical label) than individuals who are less susceptible to lexical priming.

Importantly, previous findings showing that lexical entrainment is influenced by situational factors (e.g., Brennan & Clark, 1996; Branigan et al., 2011, Van Baaren et al., 2003), taken together with our finding that lexical entrainment is stable within individuals, suggest that research on lexical entrainment could also be informative about how individual differences and situational factors interact with each other during language use. The degree of lexical entrainment may vary between individuals not only depending on their basal propensity towards entrainment and/or on features of the communicative situation, but also as a result of the interaction between each individual's disposition towards lexical entrainment and relevant situational factors. For example, lexical entrainment research can be informative about whether personality traits interact with characteristics of our conversational partner (e.g., are individuals who are more agreeable more likely to entrain to a non-prototypical community member than individuals who are less agreeable?).

In addition, the attested reliability of our instrument shows that, in principle, null findings in previous correlational studies on lexical entrainment may be indicative of true null associations, such as the null association reported by Hopkins et al. (2017) between individuals' propensity towards lexical entrainment and their perspective-taking skills. That said, there are still reasons for caution in interpreting such data. For one, while we have shown that lexical entrainment has strong test-retest reliability, the reliability of other instruments used in such studies (e.g., the mental attribution instrument used in Hopkins et al.) is not always known. In addition, it is also possible that test-retest reliability may vary across populations. We demonstrated substantial reliability in British English adults recruited and tested online, but reliability may differ for, e.g., the typically developing and autistic children studied by Hopkins and colleagues.

Going forward, we recommend that any instrument for studying lexical entrainment should be appropriately validated with the population that it will be applied to, and this applies not only for measuring reliability but also for

appropriately norming the low-frequency names that are used as stimuli. For instance, both studies reported in this paper used online participants, native speakers of British English who were users of Prolific [https://prolific.ac/]; we therefore normed our materials in a different sample drawn from the exact same population.

Although the attested substantial reliability of our task suggests that it is well-suited for correlational research, we reported additional analyses that interrogated which methodological features of our task contribute to this, and which features could be improved (see Follow-up Analyses section). In particular, we recommend the use of a range of different items to ensure independence of measurement in each trial and thus prevent measurement error from increasing throughout the task; accordingly, we encourage the incorporation of analyses to rule out trial order effects on entrainment. Moreover, we found that increasing the number of trials had a positive impact on the level of reliability of our task, although the benefits of increasing the number of trials decreased as the sample size increased. However, it is still possible that a task with a greater number of trials could provide a more precise measurement of individuals' basal tendency to lexically entrain. Given that a high number of participants in our studies scored close to zero, using a task with a greater number of trials would also allow distinguishing these participants, providing a more accurate measurement of each individual's tendency to entrain.

Strikingly, the fact that lexical entrainment is stable within individuals not only suggests that the mechanisms supporting lexical entrainment are stable when individuals believe themselves to be interacting with a remote player, but also that the way in which we make lexical choices can potentially reflect stable individual differences in how we process language. Lexical entrainment research has already suggested that language production is affected by memory, perspective-taking, and pro-sociality, among other factors (Branigan et al., 2011; Brennan & Clark, 1996; van Baaren et al., 2003). Future studies using an individual differences approach can further develop accounts of language processing, by addressing questions such as whether the degree to which individuals display lexical entrainment might be predicted by social psychological factors (e.g., perspective-taking skills), personality traits (e.g., degree of pro-sociality or agreeableness), cognitive effects (e.g., ease of

lexical access) or even by demographic variables (e.g., gender and age). Importantly, the fact that this lexical entrainment task is well-suited to both experimental and correlational research is promising for understanding how situational factors and individual differences interact during language processing, which has hitherto attracted little attention in psycholinguistic research.

Additionally, the finding that lexical entrainment is fairly stable across measurement sessions opens up the question of whether other types of entrainment, e.g., syntactic or phonetic, might be similarly stable. This is theoretically important because it is currently unclear whether entrainment is underpinned by domain-general mechanisms that might cause a person to entrain at similar rates for both lexical and grammatical stimuli, or whether different types of entrainment rely on importantly different processes; for example, lexical entrainment may be more sensitive to perspective-taking abilities than syntactic entrainment (e.g., Branigan et al., 2011). To test these issues, we need instruments that can reliably measure entrainment at various levels of linguistic structure. We have shown how this can be done for lexical entrainment; future studies should similarly focus on validating the test-retest reliability of instruments measuring linguistic entrainment at other levels of structure. To this end, and in light of the results reported in the Follow-up Analyses section, we suggest that future linguistic entrainment instruments aimed at studying individual differences should be designed to measure behavior independently in each trial (i.e., testing entrainment to different linguistic structures) and should include a number of trials as large as practically possible.

Likewise, the attested reliabilities of our instrument can have important consequences for theories of non-linguistic behavioral mimicry. During social interaction, people not only entrain to their interlocutor’s language use, but also to other non-linguistic behaviors, including body postures, gestures, facial expressions, and emotional reactions (see Chartrand & Lakin, 2013, for a review). And, although there is a general tendency to conceptualise linguistic entrainment as a kind of behavioral mimicry (e.g., van Baaren et al., 2003; Chartrand et al., 2005), it remains unclear whether linguistic and non-linguistic imitative behaviors are supported by the same constructs. Future studies on lexical entrainment could illuminate this debate by

interrogating whether the degree of lexical entrainment might be predicted by degrees of non-linguistic behavioral mimicry (e.g., mimicry of facial expressions of emotions) and their underlying constructs (e.g., measures of social competence, e.g., Mauersberger et al., 2015). For instance, are individuals who are more likely to entrain to a partner's lexical choice also more likely to mimic a partner's emotional facial expressions? If so, are both their tendencies to lexically entrain and to mimic their partner's emotional facial expression correlated with the same potential underlying mechanism (e.g., social competence)?

One potential concern that could be raised about this instrument is the degree to which it solely measures entrainment. For example, in an alternative formulation of this instrument, we might have measured entrainment as the difference between each participants' tendency to use low-frequency names when primed, and when not primed. However, although such a design might be ideal in theory, we suggest that its benefits in practice are small, and its disadvantages are serious. Its key advantage would be to account for participants' baseline tendencies to use the low-frequency labels, which could partially explain the correlations across sessions. However, our design accounted for such a tendency by using different labels across the two sessions, and by using low-frequency labels that were not systematically drawn from a particular dialect or register.

Moreover, the alternative subtractive design would have significant difficulty providing precise measures of entrainment. In particular, taking a baseline measure would a) prime participants to use higher-frequency labels for the relevant objects, thus minimizing subsequent entrainment (Branigan et al., 2011); b) reduce the number of trials available to measure entrainment, compromising statistical power; and c) could potentially increase measurement error, as each subject's score would now have two sources of error: one for measuring the baseline and one for measuring entrainment. These considerations, and the fact that entrainment is reliably seen at the group level, suggest that our single-measure instrument provides the most efficacious way of capturing this phenomenon.

In sum, we have argued that the study of individual variation in language processing requires instruments that elicit a wide range of effects between individuals

and that have high test-retest reliability. It is therefore necessary to develop tasks that meet these two criteria as a necessary precursor to testing theoretical accounts of language processing. In this paper, we have shown how this can be done for the case of lexical entrainment, a phenomenon that is informative of individual variability in how we make lexical choices in particular, and in how we process language more generally. In particular, we have shown that online naming-tasks measuring lexical entrainment can in principle be informative about factors affecting language processing. We therefore encourage the use of this instrument – adapted appropriately to the population of interest - for the study of individual differences.

## 3.7. References

Arnon, I. (2019). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01205-5

Berchtold, A. (2016). Test–retest: Agreement or reliability? *Methodological Innovations*, *9*, pp. 1-7, https://doi.org/10.1177/205979911667287

Branigan, H. P., & Messenger, K. (2016). Consistent and cumulative effects of syntactic experience in children's sentence production: Evidence for error-based implicit learning. *Cognition*, *157*, 250-256.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41-57.

Branigan, H. P., Tosi, A., & Gillespie-Smith, K. (2016). Spontaneous lexical alignment in children with an autistic spectrum disorder and their typically developing peers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1821-1831.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482-1493.

Chartrand, T. L., Maddux, W. W., & Lakin, J. L. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. *The new unconscious*, 334-361.

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual review of psychology*, *64*, 285-308.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American journal of mental deficiency*, *86*(2), 127-137

Clark, H. H. (1996). *Using language*. Cambridge university press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.

Eckert, P., & Rickford, J. R. (Eds.). (2001). *Style and sociolinguistic variation*. Cambridge University Press.

Gallois, C., & Callan, V. J. (1991). Interethnic accommodation: The role of norms. *Contexts of accommodation: Developments in applied sociolinguistics. Edited by Giles, H., Coupland, J. and Coupland, N., New York, NY: Cambridge University Press. 245-269.*

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in cognitive sciences*, *8*(1), 8-11.

Gregory, M., & Carroll, S. (2018). *Language and situation: Language varieties and their social contexts*. Routledge.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166-1186.

Hopkins, Z., Yuill, N., & Branigan, H. P. (2017). Inhibitory control and lexical alignment in children with an autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *58*(10), 1155-1165.

Horton, W. S. (2014). Individual differences in perspective taking and field-independence mediate structural persistence in dialog. *Acta psychologica*, *150*, 41-48.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*(4), 589-606.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*(2), 127-142.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *L. Pervin and O.P. John (Eds.), Handbook of personality: Theory and research (2nd ed.). New York: Guilford*, 102-138.

Juncos-Rabadán, O., Facal, D., Rodríguez, M. S., & Pereiro, A. X. (2010). Lexical knowledge and lexical retrieval in ageing: Insights from a tip-of-the-tongue (TOT) study. *Language and Cognitive Processes*, *25*(10), 1301-1334.

Kaschak, M. P., Kutta, T. J., & Jones, J. L. (2011). Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic bulletin & review*, *18*(6), 1133-1139.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in cognitive sciences*, *22*(2), 154-169.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174.

Larson, M. J., Baldwin, S. A., Good, D. A., & Fair, J. E. (2010). Temporal stability of the error-related negativity (ERN) and post-error positivity (Pe): The role of number of trials. *Psychophysiology*, *47*(6), 1167-1171.

Leaper, C., & Ayres, M. M. (2007). A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. *Personality and Social Psychology Review*, *11*(4), 328-363.

March, E., Wales, R., & Pattison, P. (2003). Language use in normal ageing and dementia of the Alzheimer type. *Clinical Psychologist*, 7(1), 44-49.

Mauersberger, H., Blaison, C., Kafetsios, K., Kessler, C. L., & Hess, U. (2015). Individual differences in emotional mimicry: Underlying traits and social consequences. *European Journal of Personality*, *29*(5), 512-529.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, *1*(1), 30.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture–word interference experiments. *Journal of memory and Language*, *35*(4), 477-496.

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, *85*(4), 935-956.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, *4*(5), 648-654.

Palomares, N. A., Giles, H., Soliz, J., & Gallois, C. (2016). Intergroup accommodation, social categories, and identities. *Giles, H. (Ed.), Communication accommodation theory: Negotiating personal relationships and social identities across contexts, Cambridge University Press,* 123-151.

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, *79*(2), 637-659.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and language*, *39*(4), 633-651.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169-190.

Polit, D. F. (2015). Assessing measurement in health: beyond reliability and validity. *International journal of nursing studies*, *52*(11), 1746-1753.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609-612.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of memory and language*, *81*, 105-120.

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. Behavior research methods, 49(2), 418-432.

Strauss, G. P., Allen, D. N., Jorgensen, M. L., & Cramer, S. L. (2005). Test-retest reliability of standard and emotional stroop tasks: an investigation of color-word and picture-word versions. *Assessment*, *12*(3), 330-337.

Van Baaren, R. B., Holland, R. W., Steenaert, B., & van Knippenberg, A. (2003). Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology*, *39*(4), 393-398.

**Note:** In sum, Chapter 3 demonstrated that our lexical entrainment task is substantially reliable for the study of individual indiferences, and that lexical entrainment thus reflects a stable individual differences in language processing. Chapter 4 uses this task to examine which mechanisms may underlie individual differences in lexical entrainment.

## Chapter 4

# 4. Individual differences in lexical entrainment

## 4.1. Introduction

It is widely recognised that different social situations cause people to vary the words they use to express themselves, such that speakers make different lexical choices when in a job interview and when chatting with friends. To do this, multiple cognitive processes are deployed, e.g., successful reference implies retrieving from memory words that appropriate for the context, and inhibiting words that are not (e.g., Levelt, 1989; Shao, Roelofs, & Meyer; 2012). But it is unclear whether and how the variability of speakers' words is determined by individual differences. We address this issue through the lens of lexical entrainment, or the tendency for a speaker to reuse a word that their conversational partner has used before (e.g., using *brolly* after their partner used *brolly*; Garrod & Anderson, 1987; Brennan & Clark, 1996; Branigan, Pickering, Pearson, McLean, & Brown, 2011). In particular, we look at whether and how individuals' tendency to lexically entrain is predicted by both age and schizotypical traits, factors that are thought to correlate with differences in how speakers access, maintain, and inhibit words during language processing (Burke & Laver, 1990; Burke & Shafto, 2004; Ettinger, Mohr, Gooding, Cohen, Rapp, Haenschel, & Park, 2015).

Research on lexical entrainment has already elucidated mechanisms involved in language production during dialogue, but there are still competing theories as to what exactly underlies the phenomenon. For instance, unmediated accounts of

entrainment argue that speakers reuse their partner's words due to recent automatic lexical processing that is independent of their beliefs about their partner: The recent processing of *brolly* made its representation accessible in memory, facilitating its retrieval and use (e.g., Pickering & Garrod, 2004; see Meyer, 1996; Meyer & Schvaneveldt, 1971; Neely, 1976). By contrast, mediated accounts of entrainment suggest that the phenomenon is mediated by speakers' beliefs about their partner. Some such accounts argue that entrainment is driven by audience design, such that speakers adopt their partner's perspective during dialogue, and thus reuse a partner's word to facilitate mutual comprehension (e.g., Brennan & Clark, 1996; Clark, 1996). Other mediated theories of lexical entrainment have suggested a prosocial component, arguing that people reuse their partner's language use to increase their social likeability (e.g., Giles & Coupland, 1991; Van Baaren, Holland, Steenaert, van Knippenberg, 2003). Evidence for accounts featuring affiliation goals comes mainly from studies showing that a speaker's likeability increases when they repeat their partner's language use, but it is still uncertain whether speakers' tendency to entrain is in fact causally linked to their desire to affiliate with their interlocutor.

Importantly, previous work has shown that lexical entrainment is affected by both ease of lexical retrieval and speakers' beliefs. Speakers tend to use a low-frequency name (e.g., *brolly*) more often after their partner has used the same name than after their partner has used a high-frequency name (e.g., *umbrella*), and their propensity to entrain increases as the lag between prime and target decreases, pointing out the role of ease of retrieval in lexical entrainment (e.g., Branigan et al., 2011; Hopkins, Yuill, & Branigan, 2017). Furthermore, speakers reuse recently processed referential expressions in a partner-specific fashion (Brennan & Clark, 1996; Horton & Gerrig, 2002, 2005), and they entrain more often to a partner who is presented as less capable versus a partner presented as more capable (e.g., Branigan et al., 2011), suggesting that entrainment is sensitive to speakers' beliefs at least in some communicative situations. In addition, speakers' tendency to repeat their partner's language use has shown to increase that partner's likeability from their addressee's perspective, suggesting a potential social affiliation component to the phenomenon (van Baaren et al., 2003). Thus, accounts highlighting lexical retrieval

and accounts featuring beliefs do not need to be mutually exclusive, and it is in turn important to understand whether and how lexical processing interacts with speakers' beliefs during entrainment.

It is worth noting that theories accounting for beliefs about the social situation do not imply the conscious strategic use of such situational information; basic memory processes, executive functions, and unconscious affiliative goals could also explain the influence of situational features on entrainment. In particular, information about the interlocutor and general situational context can act as a compound cue to retrieval of a particular name. For example, when a conversational partner (e.g., Mary) uses *brolly*, the speaker would associate the representation of Mary (i.e., MARY) to the representation of *brolly* (i.e., BROLLY), and thus the activation of the representation of Mary facilitated the retrieval and reuse of *brolly* in future interactions (Horton & Gerrig, 2005, 2016). This consideration is consistent with previous findings that executive functioning—specifically, failures in inhibitory control—can account for some occasional insensitivities to common-ground during language comprehension. In particular, individual differences in the ability to inhibit irrelevant contextual information determines the degree to which addressees successfully inhibit perspective-inappropriate interpretations of temporary referential ambiguities in their partner's speech (Brown-Schmidt, 2009). Moreover, social psychological theories of language use have suggested that speakers accommodate their language choices to their partners' language use due to an unconscious desire to affiliate with them, often without even recalling doing so (Giles & Coupland, 1991; Giles, Mulac, Bradac, & Johnson, 1987; Natale, 1973; Giles, 1973).

Critically for the study of lexical entrainment mechanisms, the tendency to reuse a partner's word varies from person to person, suggesting that lexical entrainment mechanims and/or their interactions with each other must vary across individuals (Tobar-Henríquez, Rabagliati, and Branigan, 2019). In a recent study, we measured lexical entrainment using an online interactive referential task (based on Branigan et al., 2011), where participants alternated turns with what they believed to be an online partner to either match or name a picture (in reality the 'partner' was always pre-programmed software). In this task, the experimental targets were

pictures of objects that could be named with both a disfavoured and a favoured name in the participants' speech community (e.g., *brolly* versus *umbrella*, in British English). Materials were pre-tested to ensure that participants rarely used the disfavoured name spontaneously, but still considered it an acceptable name for the object. In the main matching-and-naming task, participants always named the experimental targets after the 'partner', and lexical entrainment was then measured as the proportion of trials on which participants used the same name used by the 'partner'. In two studies we measured the test-retest reliability of lexical entrainment, both across two measurement sessions separated by minutes and across two sessions separated by 7-to-8 days. We found a considerably wide range of variation across individuals' tendency to reuse their partner's terms, and this tendency was consistent at the individual level both in the short- and the long-term, demonstrating that the phenomenon reliably reflects individual differences in language processing. However, it remains unclear what drives these differences between individuals.

One way to address this issue is understanding whether and how individuals' tendencies to entrain correlate with individual differences in potential lexical entrainment mechanisms, i.e., lexical retrieval, and audience design and social affiliation skills. Critically, such abilities undergo important within-individual changes throughout the lifespan (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Nicholas, Obler, Albert, & Goodglass, 1985), and they can also vary widely across individuals of similar age, depending on psychiatric vulnerabilities such as schizotypical traits (Beadle, Sheehan, Dahlben, & Gutchess, 2015).

As we age, we experience changes in some aspects of language production (Kemper & Anagnopoulos, 1989). In particular, and relevant to unmediated accounts of entrainment highlighting the role of lexical retrieval, older adults experience more difficulties than younger adults when retrieving words from memory. For example, they experience more tip-of-the-tongue states, take longer to find the right words for target objects (e.g., a dog), and make more naming mistakes than younger adults, using more semantically-related but inaccurate words for targets (e.g., referring to a dog as *cat*; e.g., Burke & Laver, 1990; Mitchell, 1989; Nicholas et al., 1985; Thomas, Fozard, & Waugh, 1977). These naming difficulties have been associated with

problems accessing the phonological representations of words, rather than their semantic representations. Evidence for this explanation comes from studies showing that elderly speakers make correct prime-target semantic associations (e.g., associating *dog* with *cat*) and that they successfully use phonological cues to overcome lexical retrieval difficulties (e.g., successfully recalling *dog* after being told that the word started with the letter *d*; Barresi, Nicholas, Connor, Obler, & Albert, 2000; Juncos-Rabadán, Facal, Rodríguez, & Pereiro, 2009; Kavé & Mashal, 2012; Nicholas et al., 1985). Thus, if lexical entrainment is importantly influenced by ease of lexical retrieval, then individuals' tendency to entrain should increase as age increases: Older speakers should experience more difficulties than younger speakers to retrieve a name that they have not recently processed, increasing their likelihood to entrain to their partner's (recently processed) word.

In contrast, and relevant to audience design accounts of entrainment highlighting the role of perspective-taking, healthy ageing involves a decrease in mechanisms that are potentially relevant for audience design, such as maintaining relevant information in working memory (e.g., that your partner Mary recently used *brolly*; Beni & Palladino, 2004; Logie & Morris, 2014), making new memory connections (e.g., making a connection between the representation of Mary and the representation of *brolly*; Chalfonte & Johnson, 1996; Hedden & Gabrieli, 2004), and inhibiting irrelevant information (e.g., ignoring information that is not available to Mary or that is not important for the conversation; Mayr, Spieler, Kliegl, 2001; Lustig, Hasher & Tonev, 2001). Accordingly, older adults tailor their utterances to their addressee's needs to a lesser extent than younger adults (Bortfeld et al., 2001; Healey & Grossman, 2016; Horton & Spieler, 2007; Long, Horton, Rohde, & Sorace, 2018; Lysander & Horton, 2012). For example, older adults are more likely to use ambiguous descriptions for a target than their younger counterparts (e.g., saying *the spider* when the listener can see both a small and a big spider), which suggests that speakers become less successful in considering their partner's perspective as they age (Healey & Grossman, 2016). Thus, if lexical entrainment is critically influenced by audience design, then individuals' tendency to lexically entrain should decrease as they age.

Finally, relevant to prosocial accounts of lexical entrainment, social cooperativeness increases with healthy ageing. An extensive body of research supports that older adults prioritize engaging in emotionally meaningful activities and that they are more cooperative and more emotionally empathetic than younger adults (Beadle et al., 2015; Blanchard-Fields, 2007; Carstensen, Gottman, & Levenson, 1995). In particular, social psychology studies have shown that older adults are more likely to prioritise the communal wellbeing over their personal interest; for instance, when instructed to split a given amount of money with a partner, older individuals tend to make fairer offers than younger individuals, and their economic decisions are more likely to be influenced by their partner's sadness (Beadle et al., 2015; Matsumoto, Yamagishi, Li, & Kiyonari, 2016). Consequently, if lexical entrainment is importantly affected by a prosocial component, then older adults should entrain more than younger adults.

Understanding whether and how individual differences in schizotypical traits affect entrainment can also further our knowledge of lexical entrainment mechanisms. Schizophrenia has long been related to impaired cognitive processing and bizarre language behaviours (Bleuler, 1950; Andreasen, 1986; Kuperberg, 2010a, 2010b; Rabagliati, Delaney-Busch, Snedeker, & Kuperberg, 2019), and importantly some schizotypical cognitive and linguistic characteristics are gradually distributed across the population at non-clinical levels, conforming to stable multidimensional personality traits (American Psychiatric Association, 2013; Tonelli, 2014).

Following a fairly recent shift in the conceptualization of schizophrenia's signs, current schizotypy models describe the construct as fundamentally heterogeneous, and thus encourage the development of multidimensional schizotypy measures. Although the exact number of schizotypy dimensions is not settled, current models describe positive, negative, and disorganized dimensions (e.g. Kwapil and Barrantes-Vidal, 2015; Tandon, Nasrallah, & Keshavan, 2009; Vollema & van den Bosch, 1995). In particular, the positive dimension involves disruptions in content of thought (ranging from magical ideation to delusions), perceptual oddities (including illusions and hallucinations), and suspiciousness/paranoia. The negative or deficit dimension is characterized by diminished experiences and expression, such as

poverty of speech, lack of energy, lack of motivation, inability to feel pleasure in normally pleasurable activities, and flat affect. Finally, the cognitive-behavioral disorganization dimension includes disturbances in the organization and expression of thoughts and behavior (ranging from mild deficits to formal thought disorder and severely disorganized behavior).

Relevant to unmediated accounts of entrainment featuring the role of lexical retrieval, schizotypy has been linked to increased semantic priming effects. Semantic priming describes the facilitation of the processing of a target word (e.g., *dog*) that has been presented after a semantically related prime word (e.g., *cat*), relative to a semantically unrelated prime (e.g., *computer*; Meyer & Schvaneveldt, 1971; Neely, 1971). Individuals with high schizotypy seem to activate more distal concepts than controls, suggesting faster and perhaps further spread of activation through semantic representations. For example, when asked to generate words from a semantic category (e.g., fruit names), individuals with high schizotypy offer more atypical responses, such as *cherry*, than typical ones, such as *apple* (Kiang & Kutas, 2006). Moreover, when exposed to prime-target pairs characterized by related terms (e.g., *fruit – apple*) and unrelated terms (e.g., *fruit – clamp*), people with higher schizotypy notice more semantic relations in pairs of unrelated concepts, as demonstrated by the reduction in the negativity of the N400 potential after exposure to unrelated terms (Kiang & Kutas, 2005; Kiang, Prugh, & Kutas, 2010).

Although some studies have reported semantic priming abnormalities in relation to schizotypy in general, such increased priming effects have been more consistently found in people with high scores in disorganised schizotypy compared to people with low disorganised schizotypy scores (Kiang & Kutas, 2006; Folley & Park, 2005; Prévost, Rodier, Renoult, Kwann, et al., 2010). Therefore, if lexical entrainment is importantly affected by how easy it is to retrieve a given word, then people with high disorganised schizotypy scores, who are thought to simultaneously activate more lexical labels than people with low disorganised schizotypy, should entrain less often than people with low disorganised schizotypy. Explicitly, the simultaneous activation of several lexical representations should make it difficult for

individuals with high schizotypy to retrieve the lexical representation of the recently processed word, due to lexical competition.

Relevant to audience design accounts of entrainment, a recent study showed that schizophrenic patients with disorganised symptoms reuse their partner's referential expressions less often than healthy controls, and that this group difference was mediated by individuals' Theory of Mind, or their ability to explicitly identify others' intentions and mental states (e.g., Dwyer, David, McCarthy, McKenna, & Peters, 2019). Nevertheless, previous studies have repeatedly failed to find evidence that Theory of Mind is importantly affected at non-clinical levels of schizotypy (e.g., Jahshan & Sergi, 2007), and it is thus unlikely that healthy schizotypical individuals' tendency to entrain varies due to differences in mentalising abilities.

However, consistent with decreased top-down monitoring during language comprehension in schizophrenia (Rabagliati et al., 2019), schizotypy seems to imply impairments in top-down constraints involved in audience design. In particular, individuals with high positive schizotypy exhibit impairments in cognitive mechanisms that are thought to be involved in audience design, such as inhibiting unrelated concepts, keeping relevant information in working memory, and accessing and storing new information within semantic networks (Kumari & Ettinger, 2010; Kaplan & Lubow, 2011; Park, Lenzenweger, Püschel, & Holzman, 1996; Peters, Pickering, & Hemsley, 1994; Schmidt-Hansen & Honey, 2009). In fact, healthy schizotypy has been linked to differences in production and comprehension of non-literal language, such as metaphors, irony, fauxpas, and proverbs, suggesting difficulties to adopt their addressee's perspective (Langdon & Coltheart, 2004; Rapp, Langohr, Mutschler, & Wild, 2014). Consequently, if lexical entrainment is influenced by audience design, people with high positive schizotypy should entrain less often than people with low positive schizotypy.

Finally, and relevant to prosocial accounts of entrainment, individuals with high schizotypy show reduced interest and drive to participate in social activities. In particular, individuals with high negative schizotypy report less affective reactions to neutral-, bad-, and good-valenced stimuli (Cohen, Callaway, Najolia, Larsen, Strauss, 2012), and they report decreased positive affect and pleasure in daily life, and

decreased social contact and interest (Kwapil, Brown, Silvia, Myin-Germeys, Barrantes-Vidal, 2012). Thus, if lexical entrainment is affected by prosocial motivations, then people with high negative schizotypy should entrain less often than people with low negative schizotypy.

### 4.1.1. The present study

In sum, understanding age and schizotypy differences in individuals' tendency to lexically entrain can further our knowledge of how lexical entrainment works. In this study, we examined whether and how age and schizotypy predict lexical entrainment. We conducted an internet-based study in which native speakers of British English (aged 26-60) carried out two web-based tasks in separate sessions: In the first session, they carried out an interactive online picture matching-and-naming task (Tobar-Henríquez et al., 2019); in a next-day session, they provided their age and answered an online self-reported schizotypy survey (Kwapil, Gross, Silvia, Raulin, Barrantes-Vidal, & 2017).

In the lexical entrainment task, participants alternated turns with what they believed to be an online partner to either match or name a picture (in reality the 'partner' was always pre-programmed software). Materials were normed with a new, representative internet-based sample. Given previous evidence that this task reliably elicits entrainment effects for disfavoured labels when participants have experienced the partner previously using a disfavoured label (Branigan et al., 2011; Branigan, Tosi, & Gillespie-Smith, 2016; Hopkins et al., 2017), we measured entrainment to the use of disfavoured names only: Experimental trials comprised a target that, in British English, could be named with both a highly favoured name, e.g. *umbrella*, and a disfavoured, but acceptable, name, e.g., *brolly*, and the partner always used the disfavoured name to refer to the targets. Importantly, participants always matched experimental targets (i.e., responded to their partners naming the targets) before themselves naming the targets on a subsequent trial, and we measured entrainment as the proportion of trials on which the participant used the same disfavoured name as they had previously experienced the partner using.

To measure schizotypy, we used the Multidimensional Schizotypy Scale (MSS). The MSS is a self-report survey that was designed to meet the current multidimensional theoretical model of schizotypy, measuring positive, negative, and disorganized schizotypy dimensions with separate sub-scales (Kwapil et al., 2017). Each item consists of ideas, beliefs or impressions (e.g., *I occasionally have the feeling that my thoughts are not my own*) that individuals must evaluate as true or false, and the total score for each sub-scale is the number of items endorsed in the schizotypic direction. In particular, the positive schizotypy items measure magical beliefs, referential thinking, mind reading and thought transmission, supernatural experiences, passivity experiences, unusual perceptual and somatic experiences, paranoia and suspiciousness, and special powers. Negative schizotypy items measure social disinterest, flat affect, lack of energy, poverty of speech, dimished ability to experience pleasure, and lack of motivation. Note that negative schizotypy items were designed to measure trait-like negative signs, and not simply episodic depressive symptoms (Kwapil et al., 2017). For example, such items contain the specifiers, “throughout my life...”, “I have always...”, “almost always...”, “I rarely...”, “I typically...”, “I have little or no...”. Moreover, the disorganized schizotypy items measure disorganized thought and behavior, confusion, racing thoughts, loose associations, disrupted speech, difficulty following conversations, and slowness of thought. Critically for the study of individual differences, the survey includes a wide a range of items to measure each schizotypy dimension, thus enhancing the measure’s ability to distinguish between individuals, and it has been recently shown that the measure has an excellent retest reliability (Kemp, Gross, & Kwapil, 2019).

We tested whether individuals’ degree of lexical entrainment was predicted by their age, positive schizotypy, negative schizotypy, and thought disorder. If entrainment is primarily affected by ease of lexical retrieval, then the degree of lexical entrainment should be positively predicted by age (due to increased difficulties accessing the favoured term versus the disfavoured term) and negatively predicted by positive schizotypy (because of increased lexical competition due to the simultaneous activation of multiple lexical labels). In contrast, if entrainment is affected by audience design, lexical entrainment should be negatively predicted by

age and positive schizotypy (due to impaired abilities thought to support perspective-taking). Finally, if entrainment is affected by prosociality, then the degree of lexical entrainment should be positively predicted by age (due to increased prosociality) and negatively predicted by negative schizotypy (due to decreased prosociality dispositions).

## 4.2. Method

Ethical approval for this study was obtained from the Psychology Research Ethics Committee of the University of Edinburgh (429-1718/2).

### Participants

We recruited 170 native speakers of British English, born and raised in the United Kingdom, and aged between 26 and 60 years old (M=41, SD=10). See below for details on participants' recruitment.

### Materials

We used a web-based online referential task to measure lexical entrainment (Tobar-Henríquez et al., 2019; see also Branigan et al., 2011) and the Multidimensional Schizotypy Scale to measure schizotypy (Kwapil et al., 2017).

**Lexical Entrainment Task.** In the lexical entrainment task, participants take turns with a partner to first match and then name experimental items, which comprise a target picture of an object that could be named with a favoured name (e.g., umbrella) and a disfavoured, but acceptable name (e.g., brolly) in participants' speech community.

We conducted two norming tests to adapt the materials to our population of interest. In order to create the pairs of favoured and disfavoured names for each experimental target, we conducted an initial pre-test with a different set of participants, drawn from the same population as those in the main study. 60 native speakers of British English (aged 26-60, M=40, SD=9) answered two questions in an

online survey (via Prolific). For each of 120 pictured objects, participants provided a favoured name for the picture (i.e., spontaneous naming, in response to the question: What is the first word you would use to name this object?), followed by a less-favoured name (i.e., forced naming, in response to the question: What other word could you use to name this object?). From these ratings, we gathered 50 potential target pictures, for which at least 70% of participants had provided the same favoured name, and at least 15% of participants had provided the same disfavoured name. Importantly, the disfavoured names did not consistently come from specific registers or dialects of British English. The 50 potential targets were then entered into a second rating task, in which 60 new native speakers of British English (aged 26-60, M=39, SD=10) rated the acceptability of these disfavoured names with respect to the target pictures on a scale from 1 to 7, where 1 corresponded to 'Not acceptable at all' and 7 corresponded to 'Highly acceptable'. We used this task to create the final set of 28 disfavoured names, each of which had an acceptability rating above 5.3 (M=6.1, SD=.5), and had been used with a frequency below 30% (M=7%, SD=7%). We also used the first rating task to choose 28 filler pictures, in which at least 80% of participants agreed on the same favoured name.

**Multidimensional Schizotypy Scale.** The MSS was derived and cross-validated in a sample aged 18 to 59 years old. The measure comprises statements that participants are asked to evaluate as true or false. In particular, it includes 26 positive schizotypy items (e.g., *I believe that ghosts or spirits can influence my life*), 26 negative schizotypy items (e.g., *Throughout my life I have noticed that I rarely feel strong positive or negative emotions*), and 25 disorganized schizotypy items (e.g., *Even when I have time, it is almost impossible to organize my thoughts*). Consistent with the multidimensional and spectral approach of current schizotypy models, each sub-scale is scored separately and the authors recommend no cut-off values. Importantly, the MSS was developed to have positively skewed distributions, by including items of low endorsement frequency, which reflect the rare nature of schizotypy in the general population. We therefore expect to find positively (right-

sided) skewed distributions in participants' scores for all three scales. A full list of items for each sub-scale is available at https://github.com/anitatobar/IDs.

**Procedure**

Participants were recruited to take part in this study on Prolific, using an ad available only to individuals who met our inclusion criteria (see above). The advertisement stated that participants would both play a picture matching-and-naming task with a remote player and answer an online questionnaire about beliefs and impressions. Prolific users interested in participating in the study were redirected from Prolific to a Qualtrics survey. After filling in an online consent form, they were told to wait to be matched with a remote player and, after two minutes, they were redirected to the referential task (programmed with JSPsych and available at https://github.com/anitatobar/IDs; de Leeuw, 2015), where they were asked to alternate turns with their partner to match and name one out of two pictures that would appear on the screen. On each trial, participants were asked to either wait for their partner's response so that they could select the correct (matching) picture, or to name the picture on the right/left (depending on where the target appeared, which was randomized; see Figure 4.1). Half of the trials were filler trials, on which the target picture only had a single name (e.g., *onion*). The other half were experimental trials, on which the target picture could either be named with a highly-favoured name (e.g., *umbrella*) or a less-favoured but still acceptable name (e.g., *brolly*). Thus, participants completed 28 experimental items and 28 filler items in total. The structure of the matching and naming task is illustrated in Figure 4.1. Participants alternated matching and naming trials with a 'remote player', who was in fact pre-programmed software that provided scripted answers. The trial order was fixed and the latency between matching experimental target and naming experimental target was always 3 trials, as in Figure 4.1B. Importantly, the trial structure meant that the software 'partner' always named the experimental targets before the participants, using the disfavoured names exclusively.


**Figure 4.1. A.** Examples of the matching (left) and naming (right) tasks (where the favoured word is umbrella and disfavoured is brolly). In matching trials, the participant selected the named target picture. In naming trials, they named the target. Targets were presented along with randomly selected distractors. **B.** Sequence of experimental item and filler presentation. Participants first matched an experimental target with the corresponding disfavoured name, they subsequently named a filler, matched a filler, and finally named the previously matched experimental target.

After matching and naming the 56 items, they were redirected to a second Qualtrics survey, where we checked participants' beliefs about the nature of their partner by asking *How many people did you play with during the two naming tasks?* Finally, participants where redirected to a Prolific website and received a completion

code in order for us to confirm their payment for participating in the first session. The next day, we posted a new Prolific advertisment visible only to participants who had answered the matching-and-naming task. The advertisement invited participants to answer an online survey where they had to rate statements about themselves. Participants who were interested in taking part in the second part of the study were redirected to a Qualtrics survey. In this survey, they provided their age and answered the MSS. Instructions for the MSS were as follows: *The following items inquire about a broad range of attitudes, experiences, and beliefs that people have. Please answer each item in the way that best describes you. Please note that there are no right or wrong answers – just answer in the way that is most like you*. Finally, participants were redirected to a Prolific website and received a completion code in order for us to confirm their payment for participating in the second session.

## 4.3. Results

**Data processing and exclusions.**

First, we coded all naming trials for whether they showed Lexical Entrainment (using the disfavoured name used by the partner) or not (using any other name to name the target). Occasionally, participants named or selected the distractor instead of the target; these trials were coded as NA. Participants individual lexical entrainment scores were calculated as the percentage of experimental trials on which they reused the disfavoured name their partner had used before. Second, following Kwapil et al. (2017), we scored the MSS's answers as 1 when they were in the schizotypy direction and as 0 when in the non-schizotypy direction, and we summed the scores of each sub-scale (i.e., Disorganised Schizotypy, Positive Symptoms, and Negative Symptoms). We excluded all data from participants who reported believing that they had not played with a real person during the matching-and-naming task.

**Analyses**

All analyses were carried out in the R programming language and environment (R Development Core Team, 2016). We tested the effects of Age, Disorganised Schizotypy, Positive Symptoms, and Negative Symptoms on lexical entrainment using mixed-effect logistic regressions, using lme4 package version 1.1 (Bates, Maechler, Bolker, Walker, Christensen, Singmann, Dai, Grothendieck, & Green, 2015). All our predictors were transformed to approach normality. We always used the maximal random structure justified by our design that allowed the models to reach convergence (Barr, Levy, Scheepers, & Tily, 2013). To assess the significance of all main effects we used Wald tests. Analysis scripts and data are available at https://github.com/anitatobar/IDs.

We ran three analyses. First, we tested the presence of an entrainment effect by comparing the percentages of use of the disfavoured terms in the main experiment against the percentage of use of those terms in the spontaneous naming task. Second, we conducted a descriptive analysis of schizotypy subscales. Third, we tested whether entrainment was predicted by age and schizotypy subscales, using a mixed effects model.

**Lexical Entrainment.** Overall, participants used the disfavoured word around a third of the time (37%[SD=15%] by-participants and 36%[23%] by-items). The range of participants' individual scores was fairly wide, ranging from 14% to 73%. Critically, the by-item percentages of use of disfavoured names during the matching-and-naming task were significantly higher than the percentages of use of these names during the spontaneous naming task used to norm the materials (M=7%, SD=7%, V=1, $p<.0001$), clearly suggesting the presence of an entrainment effect (see Figure 4.2).

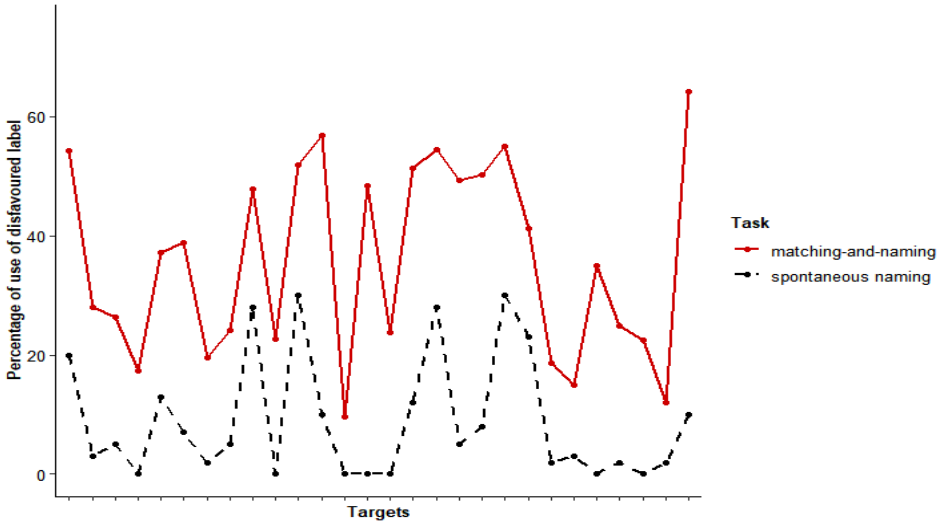


**Figure 4.2**. Percentage of use (y-axis) of disfavoured names (x-axis). The black (dashed) line represents the percentage of use of the disfavoured name in the spontaneous naming-task, while the red (solid) line represents the percentage of use of the disfavoured name in the primed matching- and naming-task used to measure lexical entrainment.

**Schizotypy Scales.** As expected, the range of schizotypy scores was fairly wide and scores for all three scales were positively (right-sided) skewed, consistent with previous results (see Table 4.1). In particular, positive schizotypy scores ranged from 0 to 23 (M=3.8, SD=4.9) and negative schizotypy scores ranged from 0 to 25 (M = 6.03, SD = 5.62). However, the range of disorganised schizotypy scores was smaller, i.e., from 0 to 15 (M = 2.95, SD = 3.87).

Table 4.1. Descriptive Statistics for Schizotypy Subscales

| | **mean** | **SD** | **median** | **min** | **max** | **skew** | **kurtosis** |
|---|---|---|---|---|---|---|---|
| *Positive Symptoms* | 3.79 | 4.91 | 2 | 0 | 23 | 1.70 | 0.46 |
| *Negative Symptoms* | 6.03 | 5.62 | 4 | 0 | 25 | 1.16 | 0.53 |
| *Thought Dis-order* | 2.95 | 3.87 | 1 | 0 | 13 | 1.23 | 0.37 |

**Individual Differences in Lexical Entrainment.** A mixed-effects logistic regression (see Table 4.2) showed that lexical entrainment was significantly and positively affected by participants' age: The older participants were, the more likely they were to entrain (β=1.24, SE=.41, z=3.1, p=.002; see Figure 4.3A). However, the model showed no evidence that the degree of lexical entrainment was affected by Disorganised Schizotypy (β=.06, SE=.1, z=.55, p>.05; see Figure 4.3B), Positive Symptoms (β=.09, SE=.14, z=.68, p>.05; see Figure 4.3C), or Negative Symptoms (β=-.05, SE=.11, z=-.44, p>.05; see Figure 4.3D).

Importantly, the correlations between predictors estimated in our model seem to be relatively small (see Table 4.3). Consistent with this, the model's estimated coefficients do not seem to be inflated as a result of multicollinearity, as suggested by the coefficients' small variance inflation values (see Table 4.4). Critically, age is not highly correlated with any schizotypy dimensions, suggesting that the positive effect of age is not likely to be inflated as a result of strong correlations with the rest of our predictors.

*Table 4.2. LexicalEntrainment ~ Age + DisorganisedSchizotypy + PositiveSchizotypy + NegativeSchizotypy+ (1|participants) + (1+Age|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -5.82 | 1.71 | -3.40 | **.000683** |
| *Age* | 1.41 | .46 | 3.07 | **.002200** |
| *DisorganisedSchizotypy* | .06 | .01 | .52 | .606500 |
| *PositiveSchizotypy* | .09 | .14 | .69 | .495700 |
| *NegativeSchizotypy* | -.05 | .11 | -.45 | .652197 |

*Table 4.3. Correlation of Fixed Effects*

| | **Intercept** | **Age** | **Disorganised Schizotypy** | **Positive Schizotypy** |
|---|---|---|---|---|
| Age | -.979 | | | |
| Disorganised Schizotypy | -.062 | .105 | | |
| Positive Schizotypy | -.206 | .100 | -.430 | |
| Negative Schizotypy | -.138 | .046 | -.383 | -.057 |

*Table 4.4. Variance Inflation Factors (VIF) of Fixed Effects*

| | **VIF** |
|---|---|
| Age | 1.06 |
| Disorganised Schizotypy | 1.62 |
| Positive Schizotypy | 1.37 |
| Negative Schizotypy | 1.30 |

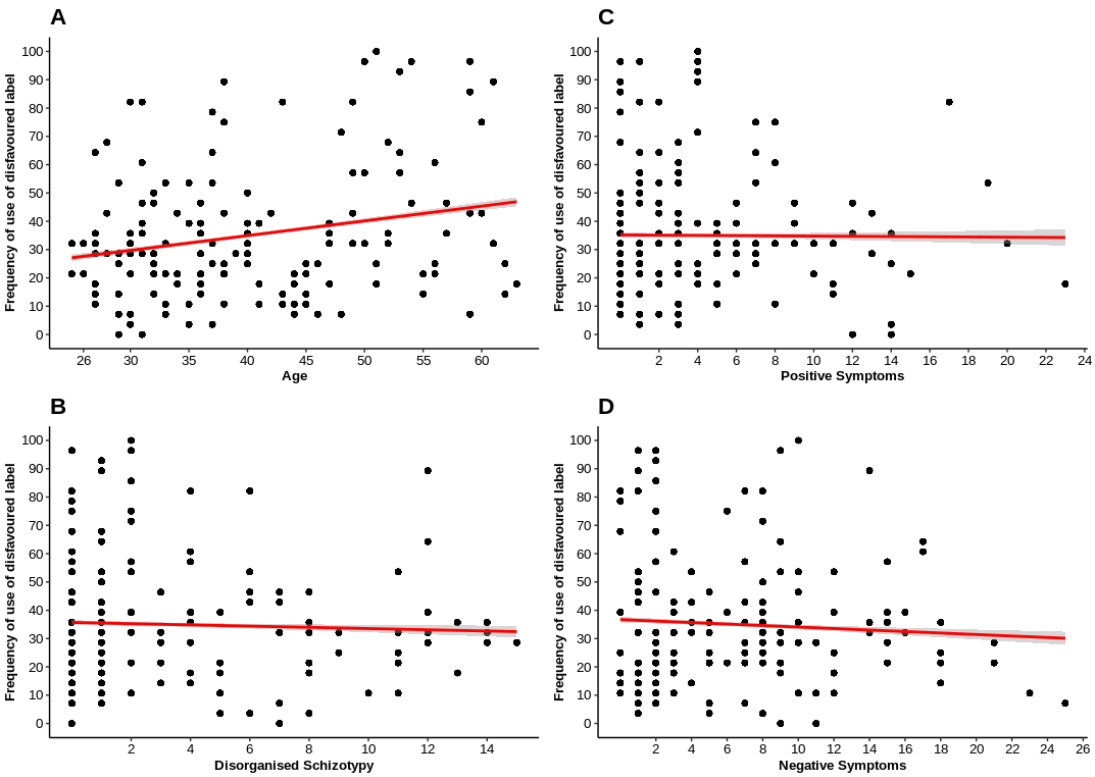

**Figure 4.3**. Correlations between participants' percentage of lexical entrainment score (y-axis) and individual differences measures (x-axis). Points are jittered. Red lines represent linear regressions between participants' total lexical entrainment scores and the relevant individual differences measure, while grey shadows correspond to a non-parametric regression smooth. **A.** Correlation between Lexical Entrainment and Age ($r$=.24, p<.0001; 95% CI [.21, .27]). **B.** Correlation between Lexical Entrainment and Disorganised Schizotypy ($r$=-.004, p>.05; 95% CI [-.03, .03]). **C.** Correlation between Lexical Entrainment and Positive Symptoms ($r$=-.009, p>.05; 95% CI [-.039, .02]). **D.** Correlation between Lexical Entrainment and Negative Symptoms ($r$=-.05, p=.001; 95% CI [-.08, -.02]).

## 4.4. Follow-up Analysis

The attested positive relationship between age and lexical entrainment might in principle reflect that at least some lexical entrainment mechanisms undergo important changes across the lifespan. However, it is also possible that our results reflect a cohort effect, so that older participants are generally more likely than younger participants to use disfavoured terms for critical items in spontaneous naming.

To examine this possibility, we tested the extent to which age predicted participants' tendency to use disfavoured terms during our (spontaneous naming) norming task. We first coded whether participants used a favoured term or a disfavoured term during the spontaneous naming question of the norming task (i.e., *What is the first question you would use to name this object?*), and then used a mixed effects model to regress the use of disfavoured names for each object against participants' age, which was transformed to approach normality.

Critically, as shown in Table 4.5, we found no evidence that participants' tendency to use disfavoured names was significantly influenced by age, supporting that the positive relationship between lexical entrainment and age was not due to a

higher baseline tendency for older participants to use disfavoured terms during spontaneous naming.

*Table 4.5. UseOfDisfavouredTerm ~ Age + (1|participants) + (1+Age|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -.71 | .98 | -.72 | .47 |
| *Age* | -.03 | .28 | -.11 | .91 |

## 4.5. Discussion

Both healthy ageing and non-clinical schizotypy traits have been linked to differences in potential lexical entrainment mechanisms: lexical retrieval, audience design, and social affiliation. To better understand the mechanisms involved in the variability of speakers' lexical choices during dialogue, we examined whether and how lexical entrainment was predicted by individual differences in age and schizotypy. In a first session, participants completed a picture-matching-and-naming task in which they alternated turns with a partner to match and name pictures with two possible names, and we measured lexical entrainment as the proportion of trials on which participants repeated the disfavoured names that their partner had previously used. In a next-day session, participants provided their age and answered a self-report mutidimensional schizotypy scale, where we measured positive, negative, and disorganised schizotypy dimensions. Although the disfavoured names had a frequency of use of 7% in a spontaneous naming-task, participants in our main experiment used disfavoured names 36% of the time, clearly suggesting the presence of a lexical entrainment effect. Importantly, the rate of lexical entrainment was not predicted by schizotypy scores, but was greater for individuals who were older.

Our results show that individual variation in the propensity to lexically entrain may not be entirely arbitrary. In particular, the fact that lexical entrainment increases with age suggests that at least some lexical entrainment mechanisms may undergo changes with healthy ageing. However, due to potential cohort effects, the

positive relationship between lexical entrainment and age needs to be interpreted with caution. For example, it is in principle possible that older participants entrained more often than younger participants due to an increased basal tendency to use disfavoured terms in everyday life. Critically, however, our follow-up analysis showed evidence against this idea: In the norming task we used, where participants spontaneously named targets and did not have the opportunity to entrain, their tendency to use disfavoured terms was not predicted by their age. This result supports that older participants did not entrain more to disfavoured terms in the main experiment because of a higher baseline tendency to use those terms during spontaneous naming.

Alternatively, it is possible that older participants in our experiment have had more experience than younger participants with the disfavoured terms used in the main experiment, and thus that their tendency to entrain was moderated by their amount of pre-experiment exposure to those terms. In particular, age could have correlated with other demographic variables, such as level of education, occupation, or socioeconomic status, which could in turn correlate with previous exposure to low-frequency terms. Our design did not control for such demographic variables; however, the disfavoured terms used in our main task did not consistently come from specific registers or regional variants of British English, and it is thus unlikely that participants' pre-experiment exposure to specific registers (e.g., formal registers) or regional variations affected their tendency to use disfavoured terms during the experiment.

With these caveats in mind, the positive relationship between age and lexical entrainment suggests that at least some lexical entrainment mechanisms might undergo changes across the lifespan. In particular, considering that ageing correlates with an increase in difficulties to access the phonological representations of words, this pattern of results may reflect older individuals' difficulties in accessing lexical representations generally, and thus their greater susceptibility to facilitation for words that have been recently processed (here, the disfavoured option brolly in preference to the normally favoured, but unprimed, option umbrella). Alternatively, since healthy ageing has also been suggested to correlate with an increased prosocial

disposition, it is possible that older individuals' increased tendency to entrain resulted from their increased desire to affiliate with their conversational partner.

Interestingly, our age findings are surprising in the light of claims that audience design skills decrease with age. If lexical entrainment was importantly affected by audience design, then older participants should have entrained less often than younger participants. Taken together with previous evidence that lexical entrainment is sensitive to audience design (Branigan et al., 2011), the positive relationship between lexical entrainment and age may suggest that audience design influences lexical entrainment only when potential communication difficulties have been made salient. For example, Branigan et al. (2011) have previously found that, in a computerised matching-and-naming task very similar to the one we used here, speakers were more likely to entrain to less capable versus more capable partners, suggesting that lexical entrainment is enhanced when successful communication is suspected to be at risk. However, potential miscommunications with the conversational partner were not made salient in our study, and it is therefore possible that audience design did not importantly affect our participants' tendency to entrain because successful communication was not suspected to be at risk. This interpretation is consistent both with previous evidence that the relevance of the addressee's perspective depends on the speakers' communicative goals (Yoon, Koh, & Brown-Schmidt, 2012), and with previous claims that lexical priming mechanisms, which automatically facilitate the use of a partner's choice, may be sufficient to yield successful communication in certain structured communicative situations (Pickering & Garrod, 2004; Hopkins et al., 2017).

The lack of evidence for lexical entrainment being affected by schizotypy dimensions can have multiple interpretations. It is possible that language deficits in healthy schizotypy are not severe enough to affect lexical entrainment. For example, given preserved functioning in other dimensions, such as Theory of Mind, healthy people with high schizotypy may develop compensatory mechanisms to exhibit functional language use during dialogue. Another possibility is that previous evidence for lexical retrieval (e.g., Kiang & Kutas, 2006), audience design (e.g., Kumari & Ettinger, 2010), and social affiliation (e.g., Kwapil et al., 2012)

impairments in schizotypy have resulted from false positive associations, perhaps due to the poor retest reliability of the schizotypy measure and/or the poor reliability of the cognitive skills measures used. Since both the lexical entrainment task and the schizotypy survey used here have been previously shown to be reliable for individual differences research (Tobar-Henríquez et al., 2019; Kemp et al., 2019), our null findings are not likely to reflect a spurious null association as a result of low retest reliabilities (Hedge et al., 2018).

Moreover, consistent with previous results from large samples, schizotypy scores were positively skewed and exhibited fairly wide ranges, which in principle would suggest that our sample is representative of the distribution of schizotypy in the general population (e.g., Kwapil et al., 2017; Kemp et al., 2019). Nevertheless, the exact values of skewness and kurtosis are also relatively lower than the values from previous studies (e.g., Kwapil et al., 2017; Kemp et al., 2019), which in contrast may suggest that our sample was not large enough to faithfully represent the distribution of schizotypy traits in the general population. All the same, what is clear from our results is that schizotypy might be a somewhat problematic measure to understand individual differences in lexical entrainment.

Importantly, however, the positive relationship between entrainment and age opens a new research direction in the study of the phenomenon, encouraging further research to tease apart the components underlying this effect. For example, if the positive age effect we found is due to retrieval difficulties and thus suggests an effect of priming on lexical entrainment, older individuals' tendency to entrain should be predicted by their difficulties in retrieving lexical representations from memory. Interestingly, young children have a high propensity to reuse referential expressions previously used by their conversational partner, even when they may be referring to a different referent from that partner. This maladaptive increased lexical entrainment is thought to result from difficulties suppressing automatic (primed) lexical activation in young children (Garrod & Clark, 1993), an idea that is consistent with evidence that individual differences in perspective-taking during comprehension are predicted by differences in individuals' inhibitory control (Brown-Schmidt, 2009; but see Hopkins, Yuill, & Branigan, 2017). It is thus possible that older adults entrained more

not only because of difficulties retrieving favored lexical labels, but also because of difficulties inhibiting recently processed (disfavored) labels. If this is the case, and the positive age effect we found results from difficulties inhibiting recent lexical information, then older individuals' tendency to entrain should be negatively predicted by inhibitory control.

In contrast, if the age effect on lexical entrainment results from social affiliation goals and thus suggests that lexical entrainment relies on prosociality, older individuals' tendency to entrain should be predicted by their prosocial inclinations. For example, under the assumption that lexical entrainment interacts with social affect and pro-sociality (e.g., Van Baaren, 2003), an older adult with a high propensity towards pro-sociality (agreeableness) should entrain more often than a person with a high propensity to feel very anxious in social situations (neuroticism). Similarly, older participants should be more sensitive than younger adults to social affiliation manipulations.

In sum, our study suggests that individual differences in lexical entrainment might be explained by mechanisms that vary across the lifespan, opening a new research direction to understand lexical entrainment components and, by extension, the variability of speakers' lexical choices. In particular, we have suggested that, as people age, the variability of their lexical choices might be informed or restricted by lexical retrieval, inhibitory control, and/or social affiliation skills.

Chapter 5

# 5. Lexical entrainment as an affiliative behaviour

## 5.1. Introduction

Being left out from a group (or ostracised) threatens individuals' sense of social belonging, which is an evolutionary determined, fundamental human need (DeWall, 2013). Consequently, experiencing ostracism motivates individuals to affiliate with others in subsequent social interactions, in order to regain social acceptance and satisfy the fundamental need to be part of a social group (Cialdini & Goldstein, 2004; Williams, Cheung, & Choi, 2000; Williams, 2007). An effective means to achieve affiliation is the reflexive mimicry of a partner's mannerisms, a behaviour that has been shown to increase individuals' social likeability and belonging (Chartrand & Bargh, 1999). Critically, when people are ostracised they tend to non-consciously mimic their partner's motor behaviours as a compensatory affiliation mechanism (Lakin, Chartrand, & Arkin, 2008). However, it is unclear whether such ostracism-induced mimicry is a fundamental means for social affiliation that extends to all aspects of behaviors, or is restricted to only certain (non-functional) behaviors. For example, do the mimicry-inducing effects of ostracism extend to language use during goal-directed social interaction? Moreover, there is no consensus about whether ostracism is a strong situation, i.e., an event that produces similar affective and behavioral responses across individuals regardless of their personalities (e.g., Monson, Hesley, & Chernick, 1982; McDonald & Donnellan, 2012), and about what potential individual variations in responses to ostracism might tell us about the nature of the underlying mechanisms and functions of mimicry.

This paper aims at examining the effects of ostracism on linguistic mimicry, and whether such effects are mediated by individual differences in personality. In two experiments, we examine whether experiencing ostracism affects individuals' tendency to imitate a conversational partner's lexical choice (*lexical entrainment*, e.g., referring to an object as *brolly* after a partner used *brolly* to name the same object; Branigan, Pickering, Pearson, McLean, & Brown, 2011; Brennan & Clark, 1996; Garrod & Anderson, 1987), and the extent to which such effects are associated with repairing specific relationships with a particular partner versus promoting affiliation more generally. We further examine whether lexical entrainment effects of ostracism, if any, are mediated by individual differences in neuroticism, a trait that has been linked to social aversion, fear of social rejection, and an increased need for social approval (John & Srivastava, 1999; Newby, Pitura, Penney, Klein, Flett, & Hewitt, 2017).

Being ostracised from a social group can inflict serious emotional distress (Baumeister & Leary, 1995; Macdonald and Leary, 2005). When people experience ostracism, they exhibit increased brain activity in regions linked to physical pain and show increased levels of the stress hormone cortisol (Eisenberger, Lieberman, & Williams, 2003; Gunnar, Sebanc, Tout, Donzella, & van Dulmen, 2003). These physiological effects complement self-report findings that ostracism decreases mood and important emotional needs, such as sense of belonging, self-esteem, and meaningful existence (e.g., Gerber & Wheeler, 2009; Zadro, Williams, & Richardson, 2004).

The emotional consequences of ostracism seem to emerge from a pervasive and automatic sensitivity to rejection, such that they occur even when people do not interact face-to-face with a partner, and when they believe that their partners have no intention to ostracise them. For example, in two experiments Zadro and colleagues (2004) had participants play a computerised ball-tossing game with three confederates who ostracised them from the game (*Cyberball*; Williams et al., 2000; Williams & Jarvis, 2006): Participants were thrown the ball only during the first two turns, and then never again. They measured the impact of experiencing ostracism on participants' satisfaction of emotional needs (e.g., sense of belonging, self-esteem,

and meaningful existence). Critically, they manipulated participants' beliefs about whether the other Cyberball players were computers or humans, and about whether these computer- or human-partners had been scripted (or explicitly told) to ostracise them. Interestingly, participants not only reported similar levels of emotional impact when interacting with computer- and human-partners, but this effect was also similar when participants believed ostracism to be intentional versus scripted.

However, previous research has suggested that, although ostracism during computer-mediated interactions has negative emotional effects, being ostracised in face-to-face interactions can have even more pervasive emotional consequences (Williams, Goven, Croker, Tynan, Cruickshank, & Lam, 2002). In particular, Williams and colleagues found that when participants were ostracised in computer-mediated interaction, they allowed themselves to exhibit acts of what the authors called 'virtual bravado', i.e., bold or intimidating verbal behaviours (e.g., *u 2 can keep talking btw yourselves and ignore me, I don't mind!!!!!!*). They suggested that the virtual bravado allowed participants to keep engaged in interaction, which might have buffered negative effects of ostracism.

Given the pervasive emotional effects of ostracism, people who have been ostracised show a wide variety of compensatory affiliation behaviours. In particular, after experiencing ostracism, people are more participative in collective group tasks, tend to conform to the opinion of others and comply with social pressure, and show increased interest in taking part in social groups (Carter-Sowell, Chen, & Williams, 2008; Maner, DeWall, Baumeister, & Schaller, 2007; Williams & Sommer, 1997; Williams et al., 2000). Such an ostracism-induced increase in affiliative behaviors even takes place when pursuing social connections could be detrimental to an individual's best interests. For example, ostracised participants are more likely than non-ostracised participants to spend money on a personally unappealing item just to please a peer, supporting the idea that post-ostracism compensatory affiliation may also be non-conscious (Mead, Baumeister, Stillman, Rawn, & Vohs, 2011).

Critically, it is not clear if ostracism-induced affiliative behaviours are targeted to repairing a particular social link (with the perpetrator of ostracism) or

enhancing affiliation in general (with any partner). For example, ostracised participants exhibit the same amount of increased affiliative behaviours with their perpetrators and 'innocent' individuals (e.g., Weerdmeester & Lange, 2019), and they pay more attention and are better at recalling social information in general, even when that information involves interactions between third parties and do not offer a direct opportunity for interaction, in turn supporting that post-ostracism increased affiliative dispositions may well be non-conscious (Gardner, Pickett, & Brewer, 2000; Hess & Pickett, 2010). But there is also evidence that ostracised individuals may retaliate against their perpetrators when possible (e.g., Twenge and Campbell, 2003; Buckley, Winkel, & Leary, 2004), which could in principle suggest that ostracism-induced increased affiliative dispositions are targeted to partners who have not perpetrated ostracism. It is worth noting that we are not aware of any previous evidence that ostracised individuals show increased affiliative behaviours to a larger extent when interacting with the same individuals who had previously ostracised them, than with an individual with whom they had not previously interacted.

Several studies have reported that imitating a social partner considerably increases an individual's social likeability, and that people can strategically engage in behavioural mimicry (albeit non-consciously) to increase their social likeability when their social belonging has been threatened by ostracism (e.g., Chartrand & Bargh, 1999; Chartrand, Maddux, & Lakin, 2005; Lakin, Jefferis, Cheng, & Chartrand, 2003; Lakin & Chartrand, 2003, 2013; Lakin et al., 2008; Cheung, Slotter, & Gardner, 2015). For example, Lakin and colleagues (2008) manipulated participants' experience of ostracism using Cyberball, and then tested whether and how ostracism influenced participants' tendency to mimic another person's physical mannerisms in a subsequent task. Participants played Cyberball with three confederates who ostracised or did not ostracise them in the game: In the ostracism condition, participants were thrown the ball only a few times at the beginning of the game, and then never again; in the control (non-ostracism) condition, participants received the ball as many times as the confederates. In the second task, participants described a set of photographs to a new confederate who had not played Cyberball; this allowed the measurement of ostracism-induced affiliation in general, and not only to the

perpetrator of ostracism. This confederate shook their foot throughout their interaction with participants, and participants' mimicry was measured as the extent to which they mimicked the confederate's foot-shaking.

The authors found evidence for a bidirectional link between imitation and affiliation, i.e., the 'perception-behaviour link' (Dijksterhuis & Bargh, 2001), such that people experience increased liking of partners who mimic their mannerisms, and tend to mimic partners with whom they want to affiliate more than those with whom they do not (Chartrand & Bargh, 1999; Stel & Vonk, 2010). Specifically, after controlling for the amount of foot-shaking that occurred during a baseline period, ostracised participants mimicked the foot-shaking behavior of the confederate significantly more than non-ostracised participants. Importantly, they did not report to have noticed this behavior, suggesting a non-conscious component. Moreover, when the confederate (who was blind to the participant's Cyberball condition) was asked to evaluate their interactions with each participant, they reported that the interactions with ostracised participants had gone more smoothly than those with non-ostracised participants.

However, the scope of ostracism-induced mimicry effects is unclear. We do not know whether such effects are restricted to non-functional behaviours, such as foot-shaking, or if they also permeate functional behaviors, such as language use during social interaction. It is uncontroversial that during social interactions people not only mimic their partner's mannerisms, but also their partner's language use at many linguistic levels. For example, speakers repeat their conversational partner's lexical choices (e.g., *brolly* versus *umbrella*), syntactic choices (e.g., passive versus active structures), their pronunciation, and even their pitch (Chartrand & Lakin, 2013; Branigan et al., 2011; Kaschack, Kutta, & Jones, 2011; Pardo, Urmanche, Wilman, & Wiener, 2017).

We here focus on linguistic mimicry at the lexical level, or lexical entrainment. Although lexical entrainment has been suggested to be a type of affiliative behavioral mimicry (e.g., van Baaren, Holland, Steenaert, & van Knippenberg, 2003; Chartrand et al., 2005; Palomares, Giles, Soliz, & Gallois,

2016), it remains unclear whether lexical entrainment and non-linguistic mimicry are actually supported by the same mechanisms. In particular, the phenomenon has been explained by various theoretical accounts. Some accounts suggest that the entrainment results mainly from linguistic processing that is not mediated by speakers' beliefs, so that a speaker entrains to a partner's use of *brolly* because the recent processing of *brolly* has made its lexical representation accessible from memory, thus enhancing its retrieval and use. In contrast, other theories of lexical entrainment argue that the phenomenon is mediated by extra-linguistic information, such as the speaker's beliefs about the interlocutor's communicative needs (Clark, 1996) and/or the speaker's pursuit of social-affective goals (van Baaren et al., 2003; Palomares, Giles, Soliz, & Gallois, 2016). Thus, understanding the effects of ostracism on lexical entrainment can in principle illuminate which components are shared between non-functional behavioral mimicry and functional linguistic mimicry at the lexical level.

Critically for the study of ostracism effects on lexical entrainment, Branigan and colleagues have developed a lexical entrainment task that has been repeatedly shown to elicit reliable effects in group-level, experimental comparisons (e.g., Branigan et al., 2011, 2016; Hopkins, Yuill, & Branigan, 2017; see Tobar-Henríquez, Rabagliati, and Branigan, 2019, for the web-based adaptation we will use here). In this task, participants take turns with a confederate to match and name pictures. Experimental targets are pictures of objects that can be named with both a disfavoured and a favoured name (e.g., *brolly* versus *umbrella*, in British English); these materials have been pre-tested to ensure that participants rarely use the disfavoured name spontaneously, but still consider it an acceptable name for the object. In the main matching-and-naming task, participants always name the experimental targets after the confederate, who always names the target with the disfavoured name, and thus lexical entrainment is measured as participants' use of the disfavoured names that their partner has used before. Importantly, Branigan and colleagues have consistently shown that this task elicits experimentally reliable entrainment effects for disfavoured names. Specifically, individuals are more likely to use a disfavoured name (e.g., *brolly*) after the partner has used that name (e.g.

*brolly*) than after the partner has used the favoured alternative (e.g., *umbrella*) or compared to the disfavoured name's baseline frequency of use (Branigan et al., 2011, 2016; Hopkins et al., 2017).

It is unclear if lexical entrainment implies a social affiliation component. Under a social affiliation account of lexical entrainment, the phenomenon can be at least partly explained by the perception-behaviour link; that is, speakers reuse their partner's referential expressions to express affiliation and enhance social relations, and this linguistic behaviour makes interactions both effective and rewarding (Van Baaren et al., 2003; Reitter & Moore, 2014; Palomares et al., 2016). However, this account is based mostly on somewhat indirect evidence. For example, speakers' tendency to repeat their partner's utterances verbatim increases their partner's affiliative behaviours (van Baaren et al., 2003), and speakers tend to entrain to their partner's language use in situations where they are generally expected to show increased affiliation, such as when interacting with a partner from a higher-status social group (Palomares et al., 2016). But it is still unclear whether lexical entrainment follows from experiences that increase affiliative dispositions, such as experiencing ostracism. Thus, the causal relationship between affiliative motivations and lexical entrainment has yet to be tested. In particular, the social affiliation account of lexical entrainment predicts that people who have demonstrably experienced ostracism should show an increased likelihood of entraining at the lexical level with a partner.

The underlying mechanisms of lexical entrainment can also be investigated by looking at how the tendency to lexically entrain varies across individuals. Importantly, lexical entrainment is reliable not only for group-comparison studies, but also reliably reflects stable individual differences (Tobar-Henríquez et al., 2019). In particular, Tobar-Henríquez et al. found a considerably wide range of variation across individuals' tendency to lexically entrain to a conversational partner, and speakers' lexical entrainment tendency was consistent at the individual level both across two sessions separated by minutes and two sessions separated by a week.

However, it is unclear what exactly underlies individual differences in lexical entrainment. An interesting possibility is that individuals' tendency to lexically entrain covaries with their social affiliation dispositions. Under the social affiliation account of lexical entrainment, a speaker's tendency to lexically entrain should not only be increased by experiencing ostracism, such that ostracised individuals should entrain more than non-ostracised individuals, but baseline social affiliation dispositions should also predict speakers' tendency to lexically entrain in general.

For example, neuroticism is a stable personality trait linked to social anxiety and high need for social approval (e.g., Eysenck and Eysenck, 1991; Newby et al., 2017). Highly neurotic people tend to be socially anxious and to seek social approval, while people with low neuroticism scores tend to be calm, even-tempered, and relaxed (Eysenck and Eysenck, 1991). If individual differences in lexical entrainment are at least partially explained by social affiliative dispositions, then we should expect to find that individual differences in lexical entrainment correlate with individual differences in neuroticism.

However, the exact relationship between entrainment and neuroticism is not clear. For example, one could argue that, given previous findings that neuroticism correlates with social anxiety and aversion (Newby et al., 2017), high-neurotic individuals should generally entrain less often than low-neurotic individuals; in contrast, given the positive relationship between neuroticism and the need for social acceptance (Eysenck and Eysenck, 1991), it could be argued that high-neurotic individuals should entrain more often than low-neurotic individuals.

It is also unclear if ostracism is a *strong situation*, that is, if the affiliative effects of ostracism vary from person to person, depending on their susceptibility to social rejection and their natural affiliative dispositions (McDonald & Donnellan, 2012). Although there are mixed results about whether personality predicts reactions to ostracism (e.g., McDonald & Donnellan, 2012; for a review, see Williams, 2007; see also Gill, Harrison, & Oberlander, 2004), recent studies suggest that this line of research merits further exploration. For example, individuals with high neuroticism and high social anxiety find it more difficult than individuals with low social anxiety

to recover from the social pain inflicted by ostracism during a Cyberball game, they experience feelings of anger after only imagining being socially excluded, and show signs of automatic social avoidance behaviours, such as decreased eye contact after social rejection (Breen & Kashdan, 2011; Heeren, Dricot, Billieux, Philippot, Grynberg, de Timary, et al., 2017; Mallott, Maner, DeWall, & Schmidt, 2009; Oaten, Williams, Jones, & Zadro, 2008). It is thus possible that ostracism stresses neurotic individuals to such a high degree that it undermines affiliative behaviours, or instead that ostracism enhances affiliative behaviours to a greater extent, as an over-compensatory measure.

### 5.1.1. The present study

To address these issues, we investigated the effects of experiencing ostracism on lexical entrainment (i.e., speakers' tendency to reuse a partner's term). In two internet-based experiments, native speakers of British English first played a ball-tossing game with two confederates, which were actually computerised agents (i.e., Cyberball; Williams et al., 2000). They then played a picture matching-and-naming task, where they alternated picture-matching and picture-naming turns with a 'partner', which was in fact software programmed to provide scripted responses. Experimental trials comprised a picture of an object (e.g., an umbrella) that could be named in British English with both a favored term, i.e., *umbrella*, and a disfavored, but acceptable, term, i.e., *brolly* (as established by a pretest). The 'partner' always named critical items before participants and using the disfavored term, and we measured lexical entrainment as the proportion of trials where participants reused the same disfavored term. Finally, participants completed a self-report personality survey, where we measured neuroticism (i.e., Big Five; John & Srivastava, 1999).

Critically, to examine the effects of ostracism on lexical entrainment, we manipulated social exclusion in Cyberball: Half of participants were assigned to an ostracism condition, where they were thrown the ball only a couple of times in the beginning, and then never again; the other half of participants were assigned to a control condition, where they received the ball as many times as the other players.

Moreover, to examine whether post-ostracism lexical entrainment effects were targeted to repairing an affiliative relationship with a particular partner or to affiliating more generally, we manipulated participants' beliefs about whether their partner in the picture-matching-and-naming task had played Cyberball with them or not. Importantly, we counterbalanced participants' gender across experimental conditions, since neuroticism has been suggested to correlate with gender (Tennant, Bebbingto, & Hurry, 1982; Swickert & Owens, 2010; Lynn & Martin, 1997).

Based on previous research, we expected that participants would lexically entrain with their partner, such that they would use the disfavored term to refer to an object more often after their partner had used it than in a spontaneous naming task used to norm our materials. But if lexical entrainment involves affiliation goals, then we should find that the extent to which participants lexically entrain will differ between social exclusion conditions. In particular, if lexical entrainment is critically affected by social affiliative goals, such that affiliation effects show up even during computerised interaction, then ostracised participants should entrain more than non-ostracised participants. Similarly, if individual differences in lexical entrainment are at least partially explained by individual differences in basal social affiliative dispositions, we would expect to find that non-ostracised participants' degree of entrainment correlates with their neuroticism scores.

Moreover, if ostracism effects reflect a general increase in social affiliative needs, instead of a targeted need to repair an affiliative relationship with a particular partner, ostracised participants should entrain as often with an individual who ostracised them as with an innocent individual; otherwise, they should entrain more with an individual who ostracised them than with an innocent individual. In addition, if ostracism effects are mediated by neuroticism, we should find a relationship between post-ostracism lexical entrainment and neuroticism; however, this could be a positive relationship (i.e., reflecting a compensatory increase in social affiliative dispositions) or a negative relationshp (i.e., reflcting an inhibition of social affiliative dispositions).

## 5.2. Experiment 1: Does lexical entrainment involve an affiliative component?

Experiment 1 investigated if lexical entrainment during an online task with a partner involved a social affiliation component, by testing the effects of experiencing ostracism on lexical entrainment and by testing the relationship between individuals' degree of lexical entrainment and individual differences in neuroticism. Moreover, Experiment 1 examined whether ostracism effects on neuroticism were targeted to repair a specific relationship with a partner or to enhance social affiliation in general, and the extent to which such ostracism effects on lexical entrainment were moderated by neuroticism scores.

Participants answered two sessions across two days. In the first session, they first played a ball-tossing game (i.e., Cyberball) with two confederates (i.e., scripted computerised agents), who either ostracised or did not ostracise them: Half of participants were thrown the ball only a few times at the beginning of the game and then never again (*Ostracism Condition*); the other half were thrown the ball as often as the confederates (*Control Condition*). We named this variable **Cyberball**. After playing Cyberball, participants played a picture-matching-and-naming task with either a partner with whom they had played Cyberball or a new partner. On each trial participants were shown two images and, while alternating turns with a 'partner', they either selected (matched) or named one of the pictures. We measured whether, on naming turns, participants reused a disfavored term that their partner had used to name the picture earlier in the study (**Lexical Entrainment**). Importantly, before starting the matching-and-naming task participants learned they would play with either a partner with whom they had previously played Cyberball (*Same Partner Condition*) or a new partner (*New Partner Condition*); we named this variable **Partner's Identity**. In a next-day session, participants answered a multidimensional personality survey, where we measured **Neuroticism**.

If Lexical Entrainment involves a social affiliation component, we should find that participants' degree of lexical entrainment is increased by experiencing ostracism, so that ostracised participants entrain more often than non-ostracised

participants. Similarly, if individual differences in lexical entrainment are explained by individual differences in basal affiliative dispositions, non-ostracised participants' degree of lexical entrainment should be predicted by individual differences in neuroticism. In particular, if neurotics' social anxiety leads to decreased social affiliation dispositions, high neuroticism ostracised participants should entrain less than low neuroticism ostracised participants; in contrast, if neurotics' need for social approval increases their affiliative dispositions, high neuroticism ostracised participants should entrain more than low neuroticism ostracised participants.

Moreover, if ostracism effects on Lexical Entrainment, if any, are targeted to repairing specific social relationships, ostracised participants should entrain more often in the Same Partner Condition than in the New Partner Condition; but if ostracism effects are instead targeted to increasing social affiliation in general, then ostracised participants should entrain to similar extents in the Same Partner Condition and in the New Partner Condition.

In addition, if ostracism effects on lexical entrainment are a strong situation, ostracised-participants' tendency to entrain should not be predicted by neuroticism; but if ostracism effects are instead moderated by individual differences in neuroticism, neuroticism should predict ostracised-participants' tendency to entrain. In particular, if neuroticism inhibits individuals from exhibiting compensatory affiliative behaviours after experiencing ostracism, high neuroticism ostracised participants should entrain less than low neuroticism ostracised participants; in contrast, if neuroticism increases ostracism-induced compensatory affiliative behaviours, high neuroticism ostracised participants should entrain more than low neuroticism ostracised participants.

### 5.2.1. Method

**Participants.** We recruited 120 online participants (60 males; M=27, SD=6.3) through the portal Prolific [https://prolific.ac/]. To be included, participants had to be native speakers of British English, born and raised in the United Kingdom, and aged 18-40 years old. Participants were paid £2 for their participation. Ethical

approval for the experiments reported below was obtained from the Psychology Research Ethics Committee of the University of Edinburgh (42-1718/4).

**Design.** We used a 2 (Cyberball: *Ostracism* versus *Control*) x 2 (Partner's Identity: *Same Partner* versus *New Partner*) between-participants factorial design; in addition, we measured Neuroticism. The dependent variable was Lexical Entrainment (i.e., participants' use of disfavored terms in the matching-and-naming task).

**Materials.** All participants answered the same tasks in a fixed order. In a first session, they played a Cyberball game (Williams et al., 2012), the picture-matching game (Tobar-Henríquez, et al., 2019), and then a final game of Cyberball (inclusion trials only; see below). In a next-day session, they answered a self-reported personality measure (i.e., Big Five questionnaire; Johnson et al., 2008).

*Ostracism manipulation.* We induced feelings of ostracism using Cyberball, a computerised ball-throwing game played with confederates (actually pre-programmed software agents; Williams et al., 2012). Following Williams et al., we explained Cyberball to participants using an information screen that laid out a 'cover story' for the game. Participants were instructed to focus on using their imagination while playing the game, rather than on winning, and they learned they would play with two remote players. The task comprised 20 trials (each lasting 200 milliseconds) in a full game. In the ostracism condition, the 'partners' were programmed to throw the ball to the participant with equal probability across the first two trials; thereafter, they threw the ball to only each other, thus leading participants to experience ostracism. In the control condition, the 'partners' were programmed to throw the ball to the participant with equal probability across all 20 trials, thus not leading participants to experience ostracism.

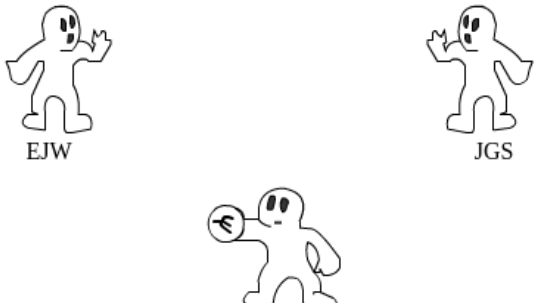


**Figure 5.1.** Cyberball trial example. Participants saw three avatars with initials below them. They controlled the avatar at the bottom of the screen, below which they saw their own initials.

Importantly, all players were represented by animated avatars on the game screen (see Figure 5.1), and each player's alleged initials were presented below their avatar, to enhance participants' belief that they were interacting with real people. We created a set of three different initials that could stand for British proper names, i.e., GHB, EJW, and JGS, which were counterbalanced across experimental conditions (see Table 5.1).

Table 5.1. Counterbalancing of partners' initials across Cyberball and Partner's Identity conditions

*Picture-matching-and-naming game.* The online picture-matching-and-naming game was identical to Tobar-Henríquez et al. (2019). The task included 15 experimental items, which comprised target pictures that could be labelled with either a favored term (e.g., *umbrella*) or a disfavored but acceptable term (e.g., *brolly*) in British English. The task also included 15 filler items, which comprised unambiguous filler pictures that could be labelled with only one favoured term (e.g., *onion*).

We conducted two norming tasks to create our experimental items. In order to create the pairs of favoured and disfavoured names for each experimental target, we conducted an initial pre-test with a different set of participants, drawn from the same

population as those in the main studies. 60 native speakers of British English (aged 18-60, M=36, SD=11) answered two questions in an online survey (via Prolific). For each of 120 pictured objects, participants provided a favoured name for the picture (i.e., spontaneous naming, *What is the first word you would use to name this object?*), followed by a less-favoured name (i.e., forced naming, *What other word could you use to name this object?*). From these ratings, we gathered 50 potential target pictures, for which at least 70% of participants had provided the same favoured name, and at least 15% of participants had provided the same disfavoured name. Importantly, the disfavoured names did not consistently come from specific registers or dialects of British English. The 50 potential targets were then entered into a second rating task, in which 60 new native speakers of British English (aged 18-60, M=38, SD=10) rated the acceptability of these disfavoured names with respect to the pictures on a scale from 1 to 7, where 1 corresponded to 'Not acceptable at all' and 7 corresponded to 'Highly acceptable'. We used this to create the final set of 15 disfavoured names, each of which had an acceptability rating above 5.3 (M=6.1, SD=.5), and had been used with a frequency below 30% (M=7%, SD=7%). We also used the first spontaneous naming task to choose 15 filler pictures, in which at least 80% of participants agreed on the same favoured name.

In each trial of the matching-and-naming-task, participants were shown two pictures (Figure 5.2), and they then either clicked on the target picture named by their partner (matching trials) or typed the name of the indicated target picture (naming trials). Half of the trials were filler trials, on which the target picture had only a single name (e.g., *onion*). The other half were experimental trials, on which the target picture could either be named with a highly-favoured name (e.g., *umbrella*) or a less-favoured but still acceptable name (e.g., *brolly*). Thus, the task used 14 experimental items and 14 filler items.

The structure of the matching and naming task is illustrated in Figure 5.1. Participants alternated matching and naming trials with a 'remote player', who was in fact pre-programmed software that provided scripted answers. The trial order was fixed and the latency between matching experimental target and naming experimental target was always 3 trials, as in Figure 5.2b. Importantly, the trial

structure meant that the software ‘partner’ always named the experimental targets before the participants, using the disfavoured names exclusively (see Figure 5.2).

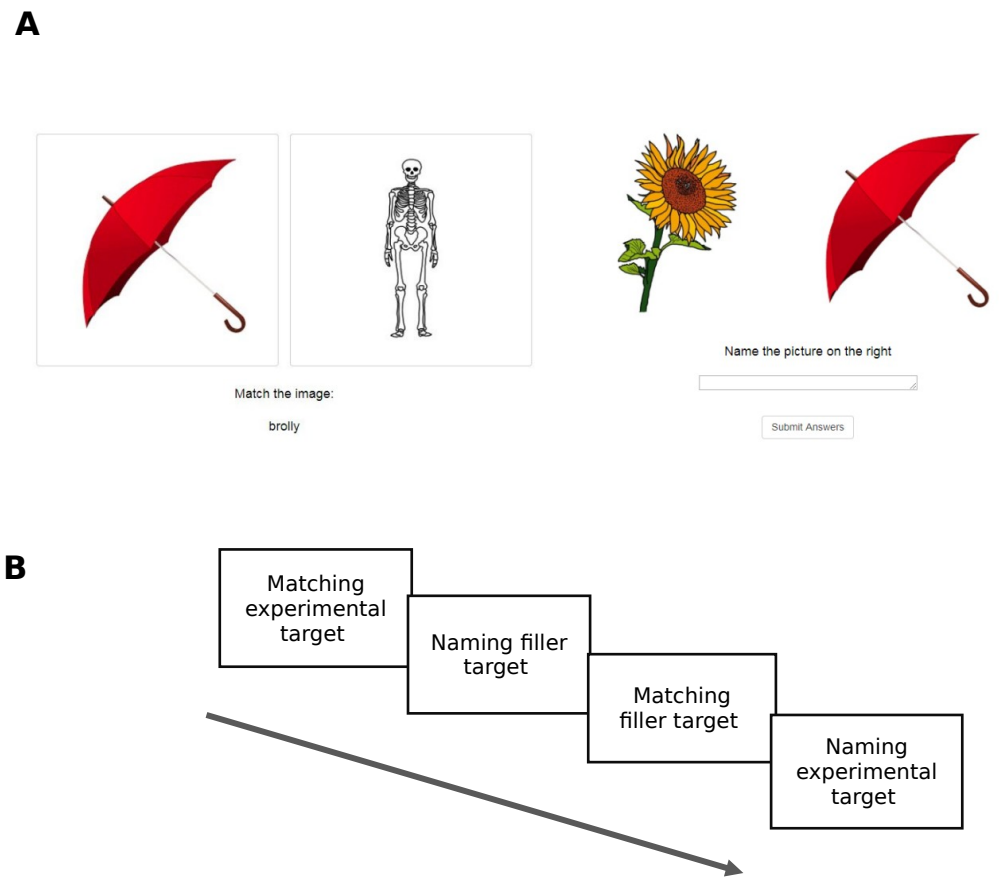


**Figure 5.2. A**. Examples of the matching (left) and naming (right) trials (where the favoured word is *umbrella* and disfavoured is *brolly*). In matching trials, the participant selected the named target picture. In naming trials, they named the target. Targets were presented along with randomly selected distractors. **B.** Sequence of experimental item and filler presentation. Participants first matched an experimental target with the corresponding disfavoured name, they subsequently named a filler, matched a filler, and finally named the previously matched experimental target.

*Personality measure.* We assessed participants’ personality traits using the Big Five survey (Johnson et al., 2008). The BFI comprises 44 items divided into 5 subscales measuring Agreeableness, Conscientiousness, Extraversion, Neuroticsm,

and Openness. Although we were interested in neuroticism only, and so we focus exclusively on neuroticism scores here, we measured all personality dimensions for comparability with previous studies. Each subscale includes a series of self-descriptive sentences (e.g. *I am someone who is talkative*), for which participants are asked to rate the extent to which they agree or disagree with that statement (from 1 to 5). These raw scores are summed to yield a score for each subscale, ranging from 0 to 40.

**Procedure.** Participants were recruited to take part in this study on Prolific, using an advertisement that was visible only to individuals who met our inclusion criteria (see above). The advertisement stated that participants would play two online games in a first session (i.e., a ball-throwing game with two partners and a picture-matching-and-naming task with one partner), and that they would answer an online survey in a next-day session. Critically, we counterbalanced participants' gender across conditions by creating advertisements available only to females or only to males for each condition. Prolific users interested in participating in the study were redirected from Prolific to a Qualtrics survey. After filling in an online consent form, they were presented the Cyberball instructions, they then were asked to type their initials for their partners to identify them during the game, and then were asked to wait to be matched with two other remote players. After a minute, they were redirected to a Cyberball game (Williams et al., 2012).

On the screen, all participants saw their partners' avatars with each partner's (alleged) initials below their avatars (see Figure 5.1); they could also see the initials that they had just typed under their own avatar. Critically, during this task we manipulated participants' experience of ostracism: Half of participants were assigned to the Ostracism Condition, where they were never thrown the ball after the first two trials; the other half were assigned to the Control Condition, where they were thrown the ball as many times as their partners. After completing 20 trials, all participants were redirected to a new Qualtrics survey, where they learned the instructions of the picture-matching-and-naming-task.

Critically, at this point we manipulated participants' beliefs about their partner. Half of participants were assigned to the Same Partner Condition, where they

were asked to wait to be connected to one of the partners with whom they had played the previous game. The other half of participants were assigned to the New Partner Condition, where they were asked to wait to be connected with a partner with whom they had not played before. After two minutes, all participants were redirected to the picture-matching-and-naming task (programmed with JSPsych and available at https://github.com/anitatobar/SocialAffectAndIDsInLexicalEntrainment; de Leeuw, 2015). Before the task started, all participants were told that their partner was waiting for them, and were asked to press the space bar to start the game. At this point they learned their partner's initials: Same Partner participants were shown one of their previous Cyberball partner's initials, while New Partner participants were shown new initials.

On each trial, participants saw two pictures and were asked to either wait for their partner's response so that they could select the correct (matching) picture, or to name the picture on the right/left (depending on where the target appeared, which was randomized) (see Figure 5.2). In matching trials, participants were given feedback on their response: They saw either a 'well done' or 'wrong answer' message when they matched the right (target) picture or the wrong (distractor) picture, respectively. In naming trials, they received feedback on their partner's matching choice, which was always positive. Half of the trials were filler trials, on which the target picture only had a single name (e.g., *onion*). The other half were experimental trials, on which the target picture could be named with either a favored term (e.g., *umbrella*) or a disfavored, but still acceptable, term (e.g., *brolly*). Importantly, the trial structure (see Figure 5.2) meant that participants always experienced their partner naming the experimental targets using the disfavoured term (e.g., *brolly*) before themselves naming the targets.

At the end of the matching-and-naming-task, all participants were redirected to a final Cyberball game where they played with two partners and were thrown the ball as many times as them. When they finished the second Cyberball game, participants were redirected to a Qualtrics survey where we ran two manipulation checks. We first checked the Cyberball manipulation, asking participants to report, in an open question, how often they were thrown the ball during the first Cyberball

game. Then, we checked participants' beliefs about the nature of their partners in both Cyberball and the matching-and-naming task, by asking how many people they had played with during each task; we coded whether participants reported playing with the right number of partners for each task or indicated that they suspected they had played with a computer. Finally, participants were redirected to a Prolific website and received a completion code in order for us to confirm their payment for their participation in the first session.

The next day, we published a new Prolific advertisement inviting participants to rate a series of statements about themselves. The advertisement was visible only to individuals who had completed the two previous online games. Participants who were interested in participating were redirected to a Qualtrics survey, where they filled a consent form and answer the Big Five survey. When finished, they were redirected to a Prolific website and received a second completion code for their participation in the second session.

### 5.2.2. Results

**Data processing and exclusions.** In the matching-and-naming task, we coded all naming trials for whether they showed Lexical Entrainment (using the disfavored term used by the partner) or not (using any other British term to name the target). We excluded the answers of 7 participants who reported believing that they had not played with a real person in the picture matching-and-naming task and of 3 participants who reported that they had not played with real people in Cyberball. From the remaining 110 participants, we exuded the answers of 10 participants who occasionally named or selected the distractor instead of the target in the matching-and-naming task; because the partner always provided a correct answer, naming or selecting a distractor without receiving any complaints from their partner could have made participants suspicious that their partner was not human. Hence our final dataset comprised data from 101 participants. Additionally, we scored the Big Five's neuroticism answers by summing the answers for each neuroticism item, following John et al. (2008).

**Analyses.** All analyses were carried out in the R programming language and environment (R Development Core Team, 2016). We tested the effects of predictors on Lexical Entrainment using mixed-effect logistic regressions, using lme4 package version 1.1 (Bates, Maechler, Bolker, Walker, Christensen, Singmann, Dai, Grothendieck, & Green, 2015)[4]. Binary dependent factors included as fixed effects were sum coded (i.e., -.5, .5), and numeric predictors were transformed to approach normality, scaled, and centred. We always used the maximal random structure justified by our design that allowed the models to reach convergence (Barr, Levy, Scheepers, & Tily, 2013). To assess the significance of all main effects and interactions involving fixed factors, we used Wald tests. We report results for key regression coefficients in the main text and full regression model results in tables; full model structures are also reported in the table captions. Moreover, for key null results we report Bayes Factors, which quantify the likelihood of observing a given data set if there were no difference across conditions, compared to if there were a difference (Wagenmakers, 2007). The analysis scripts including all models (even those that did not reach convergence) are available at https://github.com/anitatobar/SocialAffectAndIDsInLexicalEntrainment.

First, we conducted a simple comparison between ostracised and non-ostracised participants' impressions on how often they had been thrown the ball. Second, we ran descriptive analyses on neuroticism scores. Third, we conducted two analyses on participants' Lexical Entrainment. We first assessed the overall presence of an entrainment effect using a Wilcoxon test; we tested whether the produced proportion of disfavored terms was higher during the main task compared to during the spontaneous naming task used to norm the materials, where participants did not have the opportunity to entrain. Then, we tested whether entrainment was influenced by our predictors, by regressing the use of disfavored terms in each critical naming trials of the matching-and-naming-task against Cyberball (i.e., *Ostracism* versus *Control*), Partner's Identity (*Same as first game* versus *Different from first game*),

[4] Tables show models' structure using R syntax. From left to right, the first argument corresponds to the dependent variable, the second argument represents fixed effects, and arguments in parentheses are variables added as random effects (in our studies, by participants and items). In particular, random intercepts are represented with the number 1, and random slopes are represented by variables added to the right to the intercept, e.g., (1 + *slope* | items).

and their interaction. Moreover, we tested whether lexical entrainment was predicted by Neuroticism in the Control condition and in the Ostracism condition separately.

**Cyberball Manipulation-Check.** Overall, ostracised participants reported having been passed the ball less often than non-ostracised participants. Ostracised participants provided more descriptions referring to an uneven distribution of throws, e.g., ‘not very often’, ‘less than the others’, ‘almost never’, and ‘less than a third of the time’. In contrast, non-ostracised participants produced more answers referring to an equal distribution of throws between players, e.g., ‘a third of time’, ‘around a third of trials’, and ‘one time every three throws’. These results suggest that participants assigned to the ostracism condition experienced being ostracised by the confederates.

**Neuroticism.** The range of neuroticism scores was fairly wide and scores were negatively (left-sided) skewed (see Table 5.1). In particular, neuroticism scores ranged from 10 to 38 (M=25, SD=6.9).

*Table 5.2. Experiment 1: Neuroticism scores*

| | **mean** | **SD** | **median** | **min** | **max** | **skew** | **kurtosis** |
|---|---|---|---|---|---|---|---|
| **Neuroticism** | 25 | 6.87 | 25 | 10 | 38 | -.09 | -.71 |

**Lexical Entrainment Effect.** Participants used disfavoured terms on approximately a third of the trials (34.77%[SD=21%] by-participants and 34.82[18%] by-items). The by-items Wilcoxon test indicated that disfavored terms were used significantly more often in the main task than in the spontaneous naming task used to norm the materials (6%[7%]; V=0, p<.0001), clearly suggesting a Lexical Entrainment Effect (see Figure 5.3). Males (33%[19]) entrained to a similar extent as females (37%[19%]; β =-.02, SE=.24, z=-.94, p>.05; see Table 5.3).

**Ostracism effects on lexical entrainment.** Critically, ostracised participants used entrained terms (40%[24%], N=55) significantly more often than non-ostracised participants (30%[17%], N=46), suggesting that experiencing ostracism significantly increased participants’ tendency to lexically entrain in a subsequent social interaction (β =.54, SE=.24, z=2.4, p=.028; see Table 5.3 and Figure 5.3).

However, we found no evidence that ostracism effects on lexical entrainment were moderated by Partner's Identity. Overall, participants did not entrain significantly more often in the Same Partner condition (31%[19%], N=47) compared to the New Partner condition (38%[23%], N=54; β =.36, SE=.24, z=1.5, p>.05), and we found no significant interaction between Cyberball and Partner's Identity (β =.15, SE=.49, z=.32, p>.05).

*Table 5.3. Experiment 1: Lexical Entrainment ~ Cyberball*PartnersIdentity + Gender + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -.84 | .28 | -3.03 | **.00248** |
| Cyberball | .54 | .24 | 2.2 | **.02807** |
| PartnersIdentity | .36 | .24 | 1.49 | .13705 |
| Gender | -.23 | .24 | -.94 | .34981 |
| Cyberball:PartnersIdentity | .15 | .49 | .32 | .75039 |

To further explore how Ostracism and Partner's Identity might influence entrainment, we subset our data and looked at the effects of Partner's Identity on Lexical Entrainment in ostracism trials and control trials separately. As shown in Table 5.4, ostracised participants' likelihood to entrain was similar in the Same Partner condition (36%[21%], N=20) and in the New Partner condition (43%[26%], N=26; β =.43, SE=.41, z=1.1, p>.05). Similarly, and as shown in Table 5.5, non-ostracised participants' entrainment was similar in the Same Partner condition (28%[17%]; N=27) and in the New Partner condition (32%[19%], N=28; β =.28, SE=.29, z=1, p>.05).

*Table 5.4. Experiment 1 - Ostracism trials: LexicalEntrainment ~ PartnersIdentity + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -.56 | .30 | -1.9 | .0607 |
| *PartnersIdentity* | .43 | .41 | 1.1 | .2938 |

*Table 5.5. Experiment 1 - Control trials: LexicalEntrainment ~ PartnerIdentity + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -1.13 | .33 | -3.45 | **.0005** |
| *PartnersIdentity* | .28 | .29 | .96 | .3385 |

To confirm that our data supported the null hypothesis of no difference in ostracised participants' tendency to entrain to a partner with whom they had interacted in Cyberball versus a new partner, we calculated a Bayes Factor over a model assuming no difference between Partner Condition (null model) and a model assuming a difference between Partner Condition (alternative model). The null model included a fixed intercept, and random intercepts by items and participants. The alternative model included Partner Condition as main effect, and random intercepts by items and participants. We used the two models' Bayesian Information Criterion (BIC) values to estimate the Bayes Factor as $e^{(BIC\_alternative - BIC\_null)/2}$ (see Wagenmakers, 2007, and Masson, 2011). The null model fit the data slightly better, by a Bayes Factor of $e^{(1675.8-1670.7)/2} = 12.27$, with a posterior probability in favour of the null model (BF / (BF + 1) = .92), which supports that, given our data, post-ostracism lexical entrainment is more likely to be targeted to enhance social affiliation in general than to repair a specific relationship.

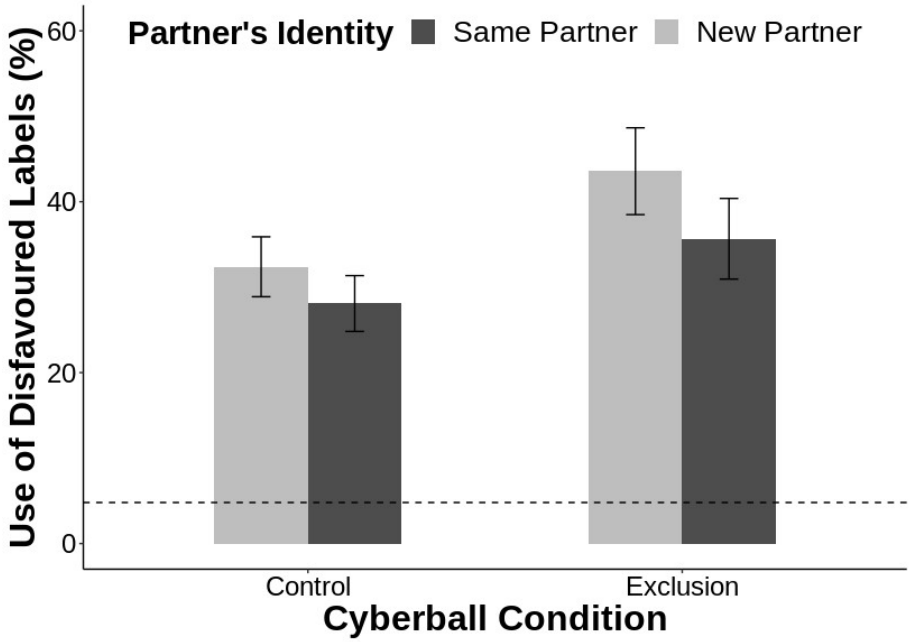


**Figure 5.3.** Mean and standard error of percentage of use of disfavoured terms (y-axis) across Cyberball Condition (x-axis) and Partner's Identity Condition (color-coded). Dashed line represents mean of percentage of use of disfavoured names in the spontaneous naming task.

**Neuroticism and ostracism effects on lexical entrainment.** As shown in Table 5.6 and Figure 5.4 (A), we found little evidence that non-ostracised participants' likelihood to entrain was moderated by neuroticism (β =-.05, SE=.14, z=-.37, p>.05). To confirm that our data supported the null hypothesis of no difference in non-ostracised participants' tendency to entrain as a function of neuroticism, we calculated a Bayes Factor over a model assuming that entrainment was not predicted by neuroticism (null model) and a model assuming that entrainment was predicted by neuroticism (alternative model). The null model included a fixed intercept, and random intercepts by items and participants. The alternative model included Neuroticism as main effect, and random intercepts by items and participants. The null model fit the data slightly better, by a Bayes Factor of $e^{(876.657-883.219)/2}$ = 26.6, with a posterior probability in favour of the null model (BF / (BF + 1) = .97). These results suggest that, given our data, it is more likely that

individual differences in lexical entrainment are not mediated by neuroticism, than that they are.

As shown in Table 5.7 and Figure 5.4 (B), we found little evidence that ostracised participants' likelihood to entrain was moderated by neuroticism (β =.19, SE=.22, z=.86, p>.05). We confirmed that our data supported the null hypothesis of no difference in ostracised participants' tendency to entrain as a function of neuroticism, by calculating a Bayes Factor over a null model including a fixed intercept, and random intercepts by items and participants, and an alternative model including Neuroticism as main effect, and random intercepts by items and participants. The null model fit the data slightly better, by a Bayes Factor of $e^{(819.36-824.8)/2}$ = 15.24, with a posterior probability in favour of the null model (BF / (BF + 1) = .94). These results suggest that, given our data, it is more likely that ostracism effects on lexical entrainment are not moderated by neuroticism, than that they are.

*Table 5.6. Experiment 1 - Control Trials: Lexical Entrainment ~ Neuroticism + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -1.14 | .33 | -3.45 | **.00056** |
| Neuroticism | -.05 | .14 | -.37 | .71219 |

*Table 5.7. Experiment 1 - Ostracism Trials: Lexical Entrainment ~ Neuroticism + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -.56 | .30 | -1.84 | .0655 |
| Neuroticism | .23 | .22 | 1.04 | .2971 |

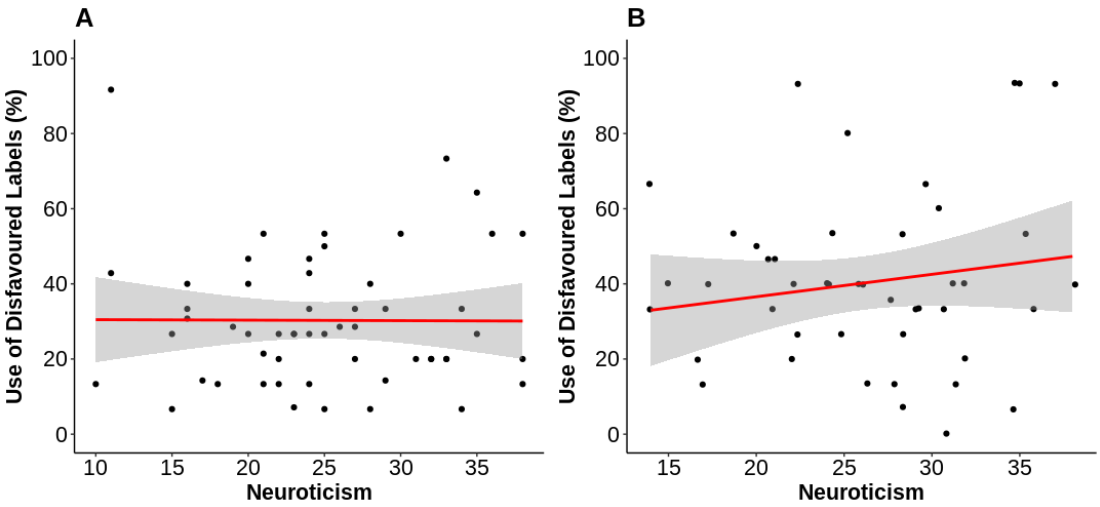


**Figure 5.4.** Experiment 1: Correlations between percentage of lexical entrainment (y-axis) and Neuroticism scores (x-axis) across Cyberball Condition. Points are jittered. The red line represents a linear regression between proportion of lexical entrainment and Neuroticism score, while the grey shadow corresponds to a non-parametric regression smooth. **A.** Correlation between non-ostracised participants' lexical entrainment and neuroticism ($r$= -.009, p>.05; 95% CI [-.27, .26]). **B.** Correlation between ostracised participants' lexical entrainment and neuroticism ($r$= .12, p>.05; 95% CI [-.17, .4]).

### 5.2.3. Discussion

Experiment 1 investigated if lexical entrainment during an online task with a partner involved a social affiliation component, by testing the effects of experiencing ostracism on lexical entrainment and by testing the relationship between individuals' degree of lexical entrainment and individual differences in neuroticism. Moreover, Experiment 1 examined whether ostracism effects on neuroticism were targeted to repair a specific relationship with a partner or to enhance social affiliation in general, and the extent to which such ostracism effects on lexical entrainment were moderated by neuroticism scores.

Overall, participants were more likely to use disfavoured terms in the main task than in a spontaneous naming task where they did not have the opportunity to

entrain, clearly suggesting the presence of a lexical entrainment effect. Critically, ostracised participants were more likely to entrain than non-ostracised participants, supporting that lexical entrainment can involve a social affiliation component. However, non-ostracised participants' degree of lexical entrainment was not moderated by individual differences in neuroticism. Importantly, ostracised participants entrained to similar extents to a partner who had just ostracised them as to a new, innocent partner, supporting that post-ostracism increased lexical entrainment is not targeted to repairing specific relationships, but rather to increasing social affiliation more generally. Moreover, we found little evidence for ostracism-induced increased lexical entrainment being moderated by individual differences in neuroticism. Experiment 2 examined the replicability of Experiment 1's findings.

## 5.3. Experiment 2: A replication

Experiment 2 investigated the replicability of Experiment 1 findings. We used the same design, materials and procedure as in Experiment 1; data processing and analyses were also identical. For scripts go to https://github.com/anitatobar/SocialAffectAndIDsInLexicalEntrainment.

### 5.3.1. Method

**Participants.** We recruited 120 online participants (60 males; M=28, SD=6) through the portal Prolific [https://prolific.ac/], following the same inclusion criteria as in Experiment 1 (see above).

**Design and Procedure.** We used the same procedure as in Experiment 1: Participants answered the same tasks in the same order, and we conducted the same experimental manipulations. Prolific users interested in participating in our study were redirected to a Qualtrics survey where they filled a consent form, and were then redirected to the ball-tossing game, where we manipulated feelings of ostracism (Cyberball: Ostracism Condition versus Control Condition). When finished, they

were redirected to the matching-and-naming task, where we measured Lexical Entrainment, and manipulated their beliefs about their partner (Partner's Identity: Same Partner versus New Partner). In a next-day session, participants who completed the first two tasks were invited to complete a self-reported survey. Those interested in participating were redirected to Qualtrics survey were we measured Neuroticism.

### 5.3.2. Results

**Data processing and exclusions.** Matching-and-naming task naming trials were coded for whether they showed Lexical Entrainment or not. We excluded the answers of 10 participants who reported believing that they had not played with a real person in the picture matching-and-naming task and of 2 participants who reported that they had not played with real people in Cyberball. From the remaining 108 participants, we excluded the answers of 7 participants who occasionally named or selected the distractor instead of the target in the matching-and-naming task. Hence our final dataset comprised data from 101 participants. Additionally, we scored the Big Five's neuroticism answers by summing the answers for each neuroticism item, following John et al. (2008).

**Ostracism Manipulation-Check.** As in Experiment 1, ostracised participants reported having been passed the ball less often than non-ostracised participants. They provided more descriptions referring to an uneven distribution of throws (e.g., 'almost never'), while non-ostracised participants produced more answers referring to an equal distribution of throws (e.g., 'as often as the others'), suggesting that participants assigned to the ostracism condition experienced being ostracised by the confederates.

**Neuroticism.** The range of neuroticism scores was fairly wide, ranging from 16 to 40 (M=29, SD=5.56), and scores were negatively (left-sided) skewed (see Table 5.8).

*Table 5.8. Experiment 2: Neuroticism scores*

| | **mean** | **SD** | **median** | **min** | **max** | **skew** | **kurtosis** |
|---|---|---|---|---|---|---|---|
| **Neuroticism** | 29 | 5.56 | 29 | 16 | 40 | -.12 | -.61 |

**Lexical Entrainment Effect.** As in Experiment 1, participants used disfavoured terms on approximately a third of the trials (39.21%[SD=26%] by-participants and 39.37[13%] by-items). The by-items Wilcoxon test indicated that disfavored terms were used significantly more often in the main task than in the spontaneous naming task used to norm the materials (6%[7%]; V=0, p<.0001), clearly suggesting a Lexical Entrainment Effect (see Figure 5.5). Males (40%[25]) entrained to a similar extent as females (38%[28%]; β =.15, SE=.30, z=-.50, p>.05; see Table 5.9).

**Ostracism effects on lexical entrainment.** As in Experiment 1, ostracised participants used entrained terms (44%[27%], N=48) around 10% more often than non-ostracised participants (35%[25%], N=53), but this difference did not reach statistical significance (β =.53, SE=.30, z=1.80, p=.076; see Table 5.9 and Figure 5.5). To confirm that our data supported the null hypothesis of no difference between ostracised and non-ostracised participants in the likelihood to entrain, we calculated a Bayes Factor over a model assuming no difference between Ostracism Conditions (null model) and a model assuming a difference between Ostracism Conditions (alternative model). The null model included a fixed intercept, and random intercepts by items and participants. The alternative model included Ostracism Condition as main effect, and random intercepts by items and participants. The null model fit the data slightly better, by a Bayes Factor of $e^{(1723.6-1728.3)/2}$ = 10.4 and a posterior probability in favour of the null model (BF / (BF + 1) = .91). Taken together, these results suggest that, given our data, it is more likely that lexical entrainment is not affected by ostracism, than that it is.

Again, we found little evidence that ostracism effects on lexical entrainment were moderated by Partner's Identity. Participants did not entrain significantly more often in the Same Partner condition (43%[28%], N=47) compared to the New Partner condition (36%[23%], N=54; β =-5., SE=.30, z=-1.70, p=.09), and there was no significant interaction between Cyberball and Partner's Identity (β =-.05, SE=.60, z=-.08, p>.05).

*Table 5.9. Experiment 2: Lexical Entrainment ~ Cyberball*PartnersIdentity + Gender + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -.57 | .24 | -2.37 | **.0180** |
| Cyberball | .53 | .30 | 1.77 | .0761 |
| PartnersIdentity | -.50 | .30 | -1.69 | .0922 |
| Gender | .15 | .30 | .50 | .6186 |
| Cyberball:PartnersIdentity | -.05 | .60 | -.08 | .9327 |

We further explored the interaction between Ostracism and Partner's Identity by looking at the effects of Partner's Identity on Lexical Entrainment in ostracism trials and control trials separately. As shown in Table 5.9, ostracised participants' likelihood to entrain was similar in the Same Partner condition (48%[30%], N=21) and in the New Partner condition (40%[22%], N=27; β=-.52, SE=.43, z=-1.19, p>.05). Similarly, and as shown in Table 5.10, non-ostracised participants' entrainment was similar in the Same Partner condition (39%[27%]; N=26) and in the New Partner condition (31%[23%], N=27; β =-.48., SE=.41, z=-1.17, p>.05).

*Table 5.10. Experiment 2 - Ostracism trials: LexicalEntrainment ~ PartnersIdentity + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -.31 | .29 | -1.08 | .280 |
| *PartnersIdentity* | -.52 | .43 | -1.19 | .235 |

*Table 5.11. Experiment 2 - Control trials: LexicalEntrainment ~ PartnerIdentity + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -.83 | .28 | -2.98 | **.00293** |
| *PartnersIdentity* | -.48 | .41 | -1.17 | .24202 |

A Bayes Factor analysis supported the null effect of partner. We compared a null model included a fixed intercept, and random intercepts by items and participants against an alternative model including Partner Condition as main effect, and random intercepts by items and participants. Again, the null model fit the data slightly better, by a Bayes Factor of $e^{(1728.59-1723.64)/2} = 11.9$, with a posterior probability in favour of the null model (BF / (BF + 1) = .92), supporting that post-ostracism increased lexical entrainment is not targeted to repairing a specific relationship (versus increasing social affiliation more generally).

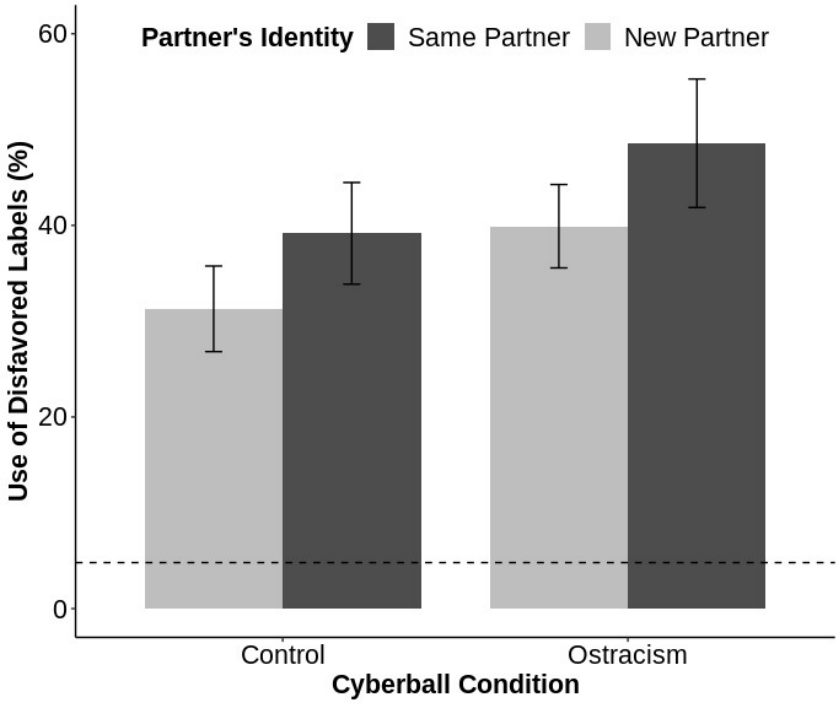


**Figure 5.5.** Mean and standard error of percentage of use of disfavoured terms (y-axis) across Cyberball Condition (x-axis) and Partner's Identity Condition (color-coded). Dashed line represents mean of percentage of use of disfavoured names in the spontaneous naming task.

**Neuroticism and ostracism effects on lexical entrainment.** As shown in Table 5.12 and Figure 5.6, we found little evidence that non-ostracised participants' likelihood to entrain was moderated by neuroticism (β =-.26, SE=.20, z=-1.34,

p>.05). A Bayes Factor analysis compared the BIC of a null model including a fixed intercept, and random intercepts by items and participants, against an alternative model including Neuroticism as a main effect, and random intercepts by items and participants. The null model fit the data slightly better, by a Bayes Factor of $e^{(897.9845-893.0673)/2}$ = 11.69, with a posterior probability in favour of the null model (BF / (BF + 1) = .92). These results suggest that, given our data, it is more likely that individual differences in lexical entrainment are not mediated by neuroticism, than that they are.

As shown in Table 5.13 and Figure 5.6, we found marginal evidence that ostracised participants' likelihood to entrain was moderated by neuroticism (β =.47, SE=.25, z=1.86, p=.06). However, a Bayes Factor analysis compared the BIC of a null model including a fixed intercept, and random intercepts by items and participants, against an alternative model including Neuroticism as a main effect, and random intercepts by items and participants; and the null model fit the data slightly better, by a Bayes Factor of $e^{(867.13-863.98)/2}$ = 4.8 and a posterior probability in favour of the null model (BF / (BF + 1) = .82). These results suggest that, given our data, it is more likely that ostracism effects on lexical entrainment are not mediated by neuroticism, than that they are.

*Table 5.12. Experiment 2 - Control Trials: Lexical Entrainment ~ Neuroticism + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -.84 | .28 | -2.99 | **.00282** |
| Neuroticism | -.26 | .20 | -1.34 | .17994 |

*Table 5.13. Experiment 2 - Ostracism Trials: Lexical Entrainment ~ Neuroticism + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -.34 | .29 | -1.21 | .2266 |
| Neuroticism | .47 | .25 | 1.86 | .06 |

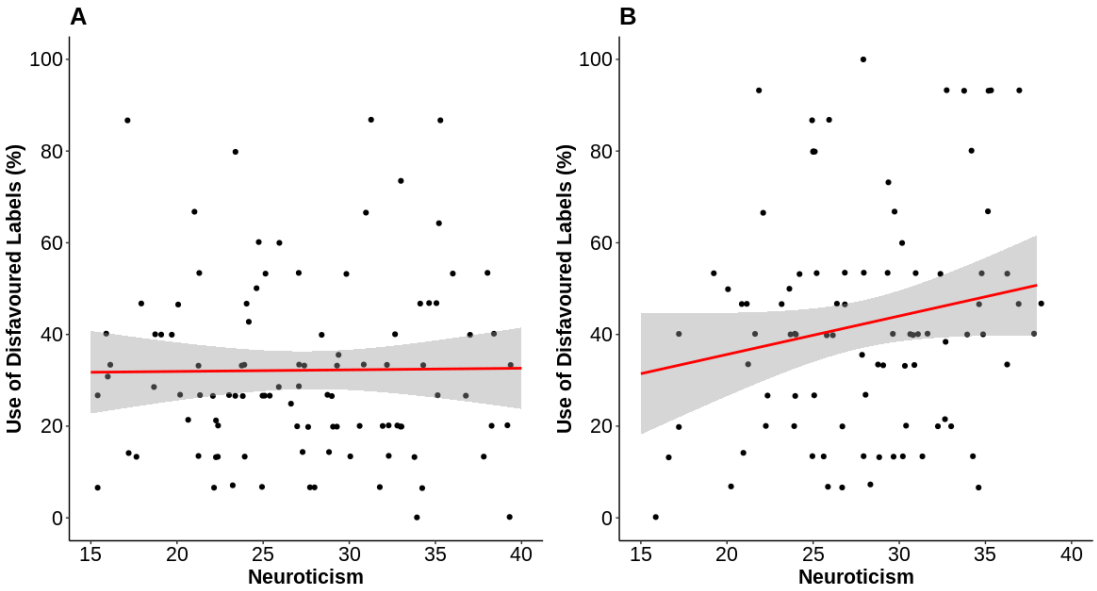


**Figure 5.6.** Correlations between percentage of lexical entrainment (y-axis) and Neuroticism scores (x-axis) across Cyberball Condition. Points are jittered. The red line represents a linear regression between proportion of lexical entrainment and Neuroticism score, while the grey shadow corresponds to a non-parametric regression smooth. **A.** Correlation between non-ostracised participants' lexical entrainment and neuroticism (*r*= -.25, p>.05; 95% CI [-.49, .02]). **B.** Correlation between ostracised participants' lexical entrainment and neuroticism (*r*= .43, **p=.002**; 95% CI [.17, .63]).

### 5.3.3. Discussion

Experiment 2 tested the replicability of Experiment 1's findings: First, that lexical entrainment is sensitive to ostracism at the situational level but that individual differences in entrainment are not predicted by neuroticism; and second, that ostracism effects on lexical entrainment are targeted to enhance social affiliation in general (versus to repair a specific relationship) and that they are not mediated by individual differences in neuroticism. As in Experiment 1, Experiment 2 showed a lexical entrainment effect: Participants were more likely to use disfavoured terms in the main experiment versus a spontaneous naming task where they did not have the

opportunity to entrain. However, although there was a numerical trend in the same direction as Experiment 1, Experiment 2 did not find a significant effect of ostracism on the likelihood of entrainment: Ostracised participants did not entrain significantly more often than non-ostracised participants. Importantly, however, Experiment 2 replicated Experiment 1's null effects regarding individual differences in non-ostracised participants' lexical entrainment, so that their tendency to entrain was not mediated by neuroticism. Experiment 2 also replicated null effects of partner: after experiencing ostracism, participants were as likely to entrain to a partner who had ostracised them versus a new partner. Nevertheless, unlike Experiment 1, Experiment 2 showed moderate evidence for the effects of ostracism being moderated by individual differences in neuroticism.

## 5.4. Follow-up Analyses

Our experiments suggest moderate evidence for an effect of ostracism on lexical entrainment. While Experiment 1 showed that ostracised participants entrained more often than non-ostracised participants, this tendency, though of a similar numerical magnitude, did not reach statistical significance in Experiment 2. In this section, we explore which methodological features of our studies may have contributed to these mixed results.

First, since the emotional effects of ostracism are thought to be transient, it is possible that ostracism effects on lexical entrainment might have washed away throughout the task, particularly given that the picture-matching-and-naming task was interactive, and so might in itself have led to social reintegration. Critically, this could have happened to different extents in each experiment, given that we used online recruitment and thus could not highly control the degree of attention that participants paid to the task in each experiment. To examine whether there was a difference in the effect of trial order in Experiment 1 and Experiment 2, we carried out a mixed model analysis regressing participants' use of disfavoured labels against the interaction between trial order and experiment. It is important to note that in our experiments trial order correlated with items: All participants encountered items in

the same random, but fixed, order, and it is thus possible that our trial order analysis is confounded by potential differences between items. That said, if the effects of ostracism washed away throughout the task, ostracised participants' tendency to entrain should decrease as trial number increased. More importantly, if our mixed results occurred due to a difference across experiments, then we should see a difference in how lexical entrainment correlated with trial number throughout Experiment 1 and Experiment 2.

Second, and perhaps more importantly, our data processing resulted in different numbers of participants in each condition for each of our two experiments. It is thus likely that the statistical power of our group-comparison analyses has varied across experiments, leading to mixed results. In principle, we could address this issue through power analyses, by estimating the appropriate sample size for a given size effect. However, since the effect of ostracism on the likelihood of lexical entrainment is not previously reported in the literature, this is not a straightforward option. Alternatively, we could base our power analysis on a size effect previously established in the non-linguistic behavioural mimicry literature. But since it is not clear whether non-linguistic and linguistic behavioural mimicry are indeed supported by the same mechanisms, it is by extension unclear to what degree the size of ostracism effects on non-linguistic mimicry should generalise to lexical entrainment. In the light of these considerations, we therefore address this issue by conducting a combined analysis of Experiments 1 and 2, thereby increasing our sample size up to 202 participants.

In addition, a potential lack of statistical power could have also impacted on our individual differences analysis, and we thus examine the extent to which neuroticism predicted post-ostracism lexical entrainment over the answers from both experiments.

**Trial order.** As shown in Table 5.14, our regression analysis indicated a significant negative effect of Trial Order on lexical entrainment across Experiments 1 and 2 ($\beta$=-.17, SE=.07, z=-2.25, **p=.025**), suggesting that participants' likelihood to entrain decreased throughout the task. Critically, however, there was no significant interaction between Trial Order and Experiment ($\beta$=-.12, SE=.08, z=1.4, p>.05),

suggesting that participants were as likely to decrease their tendency to lexically entrain during the task in Experiment 1 as in Experiment 2.

In addition, there was no significant interaction between Trial Order and Cyberball Condition (β=-.05, SE=.08, z=-.61, p>.05), suggesting that ostracised and non-ostracised participants decreased their tendency to entrain to similar extents throughout the task, in turn suggesting that the effects of trial order on entrainment is not likely to result from Ostracism effects washing away during the task. Instead, these effects seem to occur due to features of the task itself. As shown in Figure 5.7, participants' tendency to entrain does not seem to have decreased constantly throughout the task. In particular, the percentage of use of disfavoured labels varied across trials, but this variation does not seem to be related to task progress; instead, it seems to be linked to individual items. This suggests that to fully understand a potential effect of trial order it will be necessary to control for item order by counterbalancing sets of items across groups of participants, or using a randomised order of presentation. This will allow us to understand better in the future why the extent of lexical entrainment might decrease in this type of situation.

*Table 5.14. Experiments 1 and 2: Lexical Entrainment ~ TrialOrder*Experiment + TrialOrder*Cyberball + (1|participants)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -.48 | .15 | -3.24 | **.00121** |
| TrialOrder | -.17 | .07 | -2.25 | **.02476** |
| Experiment | .20 | .17 | 1.17 | .24350 |
| Cyberball | -.47 | .17 | -2.78 | **.00550** |
| TrialOrder:Experiment | .12 | .08 | 1.38 | .16908 |
| TrialOrder:Cyberball | -.05 | .08 | -.612 | .54086 |

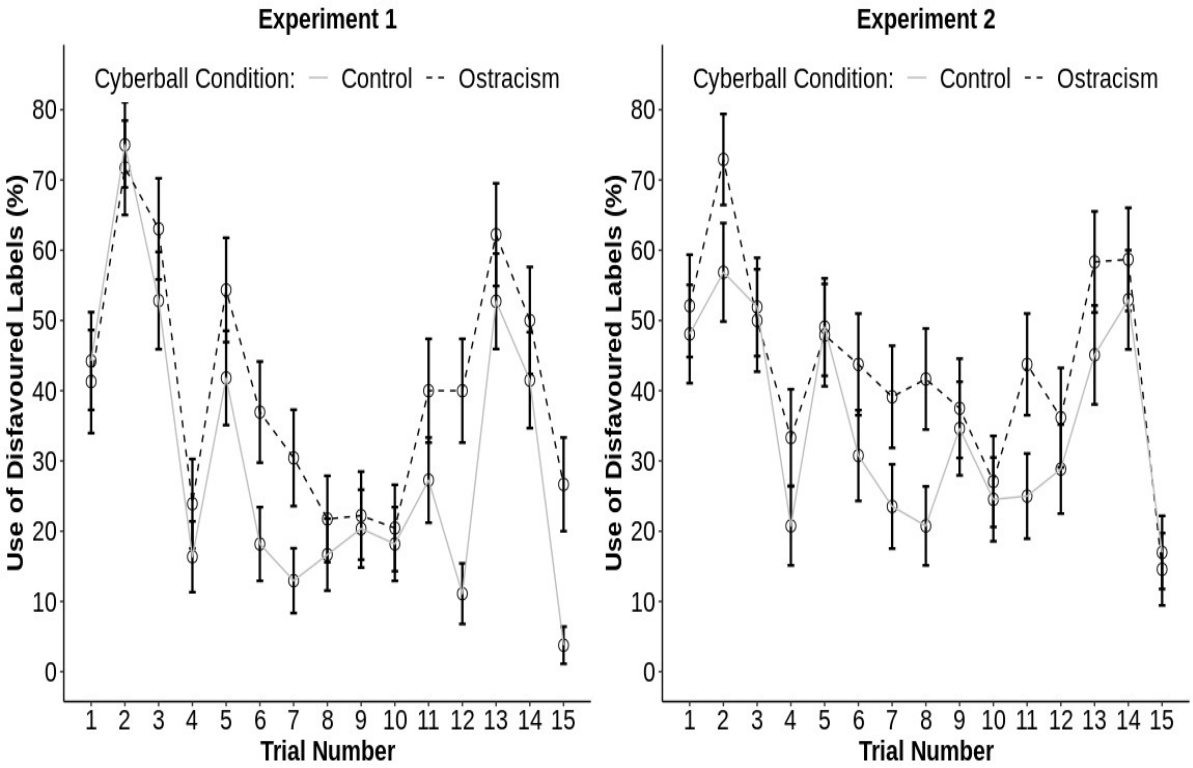


**Figure 5.7.** Mean and standard error of percentage of use of disfavoured labels (y-axis) across trial number (x-axis) and Cyberball Conditions (line-type-coded) in Experiment 1 (left) and Experiment 2 (right).

**Ostracism effects on lexical entrainment in Experiment 1 and Experiment 2.** Overall, participants entrained on 37%[24%] of naming trials (37%[15%] by-items). As expected, this tendency was significantly higher than the tendency to use disfavoured labels in the norming task (6%[7%];V=0, p<.001), and we again found no effect of gender on lexical entrainment: Females (37%[25%], N=103) were as likely to entrain as males (37%[22%], N=99; β=.008, SE=.04, z=.19, p>.05, see Table 5.12).

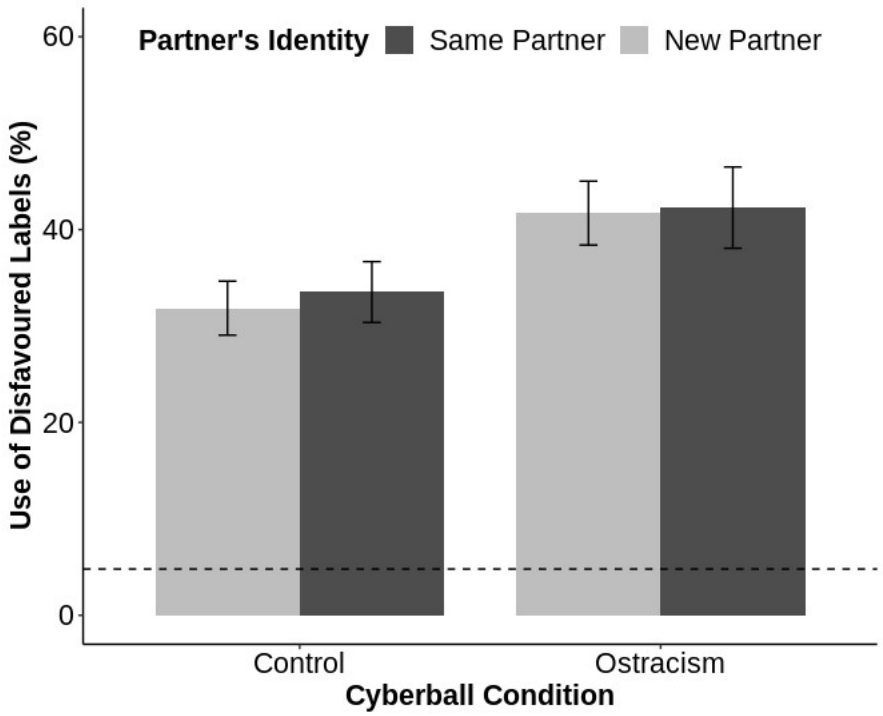


**Figure 5.8.** Mean and standard error of percentage of use of disfavoured terms (y-axis) across Cyberball Condition (x-axis) and Partner's Identity (color-coded). Dashed line represents mean of percentage of use of disfavoured names in the spontaneous naming task.

Critically, ostracised participants entrained significantly more often (42%[24%], N=94) than non-ostracised participants (32%[22%], N=108; β =.53, SE=.20, z=2.69, p=.007; see Table 5.15 and Figure 5.8), suggesting that experiencing ostracism increased the likelihood of entrainment. Moreover, the degree of lexical entrainment was similar across experiments (β=.22, SE=.20, z=1.13, p>.05), and the effects of ostracism did not seem to vary across experiments, as suggested by a non-significant interaction between Cyberball Condition and Experiment (β=-.06, SE=.39, z=-.15, p>.05). In addition, a Bayes Factor analysis compared the BIC of a null model including a fixed intercept, and random intercepts by items and participants, against an alternative model including Experiment as main effect, and random intercepts by items and participants. The null model fit the data better, by a

Bayes Factor of $e^{(3362.065-3355.427)/2}$ = 27.63 and a posterior probability in favour of the null model (BF / (BF + 1) = .96), supporting that the likelihood to entrain did not significantly vary across experiments.

As expected, participants entrained to similar extents in the Same Partner (37%[25%], N=94) and the New Partner condition (37%[23%], N=108; β=-.072, SE=.20, z=-.37, p>.05), and the effect of Cyberball was not qualified by Partner's Identity (β=.05, SE=.39, z=.12, p>.05), supporting that ostracism effects on lexical entrainment were not targeted to repairing a particular social relationship, but to increasing social affiliation more generally.

*Table 5.15. Experiments 1 and 2: Lexical Entrainment ~ Cyberball*PartnersIdentity + Cyberball*Experiment + Gender (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -.68 | .27 | -2.53 | **.01128** |
| PartnersIdentity | -.07 | .20 | -.37 | .71334 |
| Cyberball | .52 | .20 | 2.69 | **.00717** |
| Experiment | .22 | .19 | 1.13 | .25765 |
| Gender | .01 | .04 | .19 | .84668 |
| PartnersIdentity:Cyberball | .05 | .39 | .12 | .90442 |
| Cyberball:Experiment | -.06 | .39 | -.15 | .87767 |

**Neuroticism effects on post-ostracism lexical entrainment.** As expected, the range of neuroticism scores was fairly wide, ranging from 10 to 40 (M=27, SD=6.52), and scores were negatively (left-sided) skewed. However, as shown in Table 5.16 and Figure 5.9, we found no evidence that non-ostracised participants' lexical entrainment was moderated by neuroticism (β=-.33, SE=.44, z=-.75, p>.05). This was supported by a Bayes Factor analysis: a null model including a fixed intercept, and random intercepts by items and participants, fit the data better than an alternative model including Neuroticism as a main effect, and random intercepts by items and participants (Bayes Factor of $e^{(1742.514-1749.329)/2}$ = 30.20), with a posterior probability in favour of the null model (BF / (BF + 1) = .97).

However, as shown in Table 5.17 and Figure 5.9, we found moderate evidence that ostracised participants' tendency to entrain was moderated by neuroticism (β=1.26, SE=.63, z=1.99, **p=.047**). Given the p-value for this effect, we ran a Bayes Factor analysis. We compared the BIC of a null model including a fixed intercept, and random intercepts by items and participants, against an alternative model including Neuroticism as a main effect, and random intercepts by items and participants. The null model fit the data slightly better, by a Bayes Factor of $e^{(1655.12-1651.786)/2}$ = 5.29 and a posterior probability in favour of the null model (BF / (BF + 1) = .84), supporting that, given our data, it is more likely that ostracised participants' likelihood to entrain was not moderated by neuroticism, than that it was.

*Table 5.16. Experiments 1 and 2 - Control Trials: Lexical Entrainment ~ Neuroticism + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | .1 | 1.47 | .07 | .946 |
| Neuroticism | -.33 | .44 | -.75 | .454 |

*Table 5.17. Experiments 1 and 2 - Ostracism Trials: Lexical Entrainment ~ Neuroticism + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | -4.58 | 2.10 | -2.18 | **.0292** |
| Neuroticism | 1.26 | 0.63 | 1.99 | **.0468** |

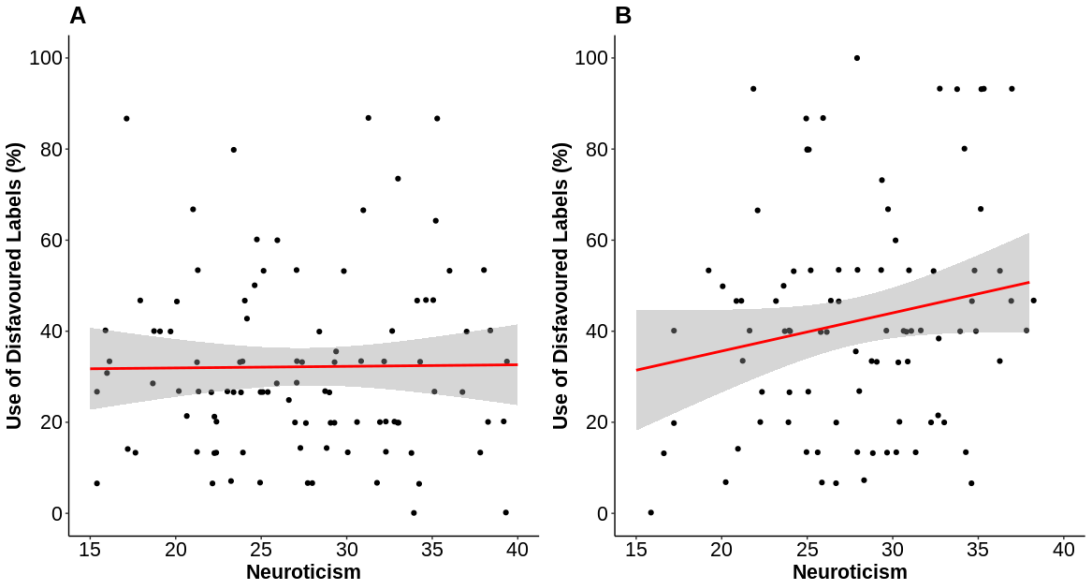


**Figure 5.9.** Correlations between percentage of lexical entrainment (y-axis) and Neuroticism scores (x-axis) across Cyberball Condition. Points are jittered. The red line represents a linear regression between proportion of lexical entrainment and Neuroticism score, while the grey shadow corresponds to a non-parametric regression smooth. **A.** Correlation between non-ostracised participants' lexical entrainment and neuroticsim ($r$= -.13, p>.05; 95% CI [-.31, .06]). **B.** Correlation between ostracised participants' lexical entrainment and neuroticism ($r$= .24, **p=.019**; 95% CI [.04, .42]).

## 5.5. Discussion

Previous research has shown that experiencing ostracism increases non-functional affiliative behaviours, such as non-linguistic behavioural mimicry (e.g., foot-shaking). However, the scope of these effects is uncertain, and we do not know whether ostracism-induced increased affiliative behaviours extend to functional behaviours, such as language use during social interaction. It is also uncertain whether ostracism-induced increased affiliative behaviours are moderated by individual dif-

ferences in personality, such as neuroticism, and whether they are targeted to repairing particular social relationships (e.g., with the person who perpetrated ostracism) or to increasing social affiliation in general (e.g., with any social partner). These are critical questions to understand which mechanisms underlie imitative behaviours, and their pervasiveness. In two experiments, we addressed these issues by understanding ostracism effects on lexical entrainment. In particular, we looked at the extent to which lexical entrainment involved a social affiliative component, both by testing (i) whether ostracism increased the likelihood of lexical entrainment at the group-level, and (ii) whether individual differences in lexical entrainment were explained by individual differences in neuroticism. Moreover, we examined whether ostracism effects on lexical entrainment with a partner were moderated by i) whether the conversational partner had ostracised the participant, and by (ii) individual differences in neuroticism.

In two experiments, participants completed three tasks in two sessions. In the first session, they first played a ball-tossing game with two confederates who either ostracised them from the game or did not ostracised them. They then played a matching-and-naming task where they took turns with a partner to match and name objects that had both a favoured or a disfavoured label (e.g., *umbrella* versus *brolly*), and we measured their tendency to reuse the same disfavoured labels as their partner had used before; importantly, the partner was either a partner from the previous game or a new partner. In a next-day session, participants answered an online survey where we measured neuroticism.

In Experiment 1, participants showed significant lexical entrainment to disfavoured labels, and the tendency to entrain was greater in ostracised participants compared to non-ostracised participants. But the post-ostracism degree of lexical entrainment was similar when ostracised participants interacted with a partner who had ostracised them as with a new, innocent partner. Moreover, ostracised participants' tendency to entrain was not significantly moderated by individual differences in neuroticism. Experiment 2, which replicated Experiment 1, also showed significant entrainment to disfavoured labels, and ostracised participants tended to entrain more than non-ostracised participants, but this difference – though numerically similar to

that found in Experiment 1 - did not reach statistical significance. Moreover, Experiment 2 replicated two of Experiment 1's null effects: Ostracised participants' lexical entrainment was similar when interacting with a partner who had ostracised them as with a new partner, and non-otracised participants' tendency to entrain was not moderated by individual differences in neuroticism. Nevertheless, it did not replicate the null effects of neuroticism on post-ostracism lexical entrainment: There was a significant positive correlation between neuroticism and ostracised participants' tendency to entrain, and the regression analysis showed a marginal effect of neuroticism in ostracism trials.

Importantly, a combined analysis of Experiments 1 and 2 suggested that these mixed results may reflect different degrees of statistical power across our two experiments, in turn supporting an effect of ostracism on lexical entrainment. In particular, our analyses included uneven numbers of participants per experimental condition across experiments, which may have caused statistical power differences. The combined analysis of our experiments (N=202) indicated that, overall, ostracised participants were more likely to entrain than non-ostracised participants, supporting that experiencing ostracism increases individuals' likelihood to entrain. This analysis also indicated that ostracised participants entrained as often to a partner who had ostracised them as to a new partner, suggesting that ostracism effects on lexical entrainment do not seem to be targeted to repairing particular relationships, but rather to increasing affiliation in general. In addition, we again found a significant correlation between neuroticism and post-ostracism lexical entrainment, and our regression analysis showed a significant relationship between neuroticism and lexical entrainment in ostracism trials, suggesting that individuals' post-ostracism degree of lexical entrainment may moderated by individual differences in neuroticism.

Critically, the fact that ostracised participants entrained more often than non-ostracised participants suggests that lexical entrainment is sensitive to ostracism. This in turn suggests that the emotional and psychological effects of experiencing social rejection from a social group are not restricted to non-functional behaviours, but they permeate functional behaviours as well, such as how individuals make referential choices during social interaction. Importantly, the finding that ostracism affected lex-

ical entrainment even during an online task, where speakers did not interact face-to-face with a partner, supports previous findings that experiencing ostracism can have pervasive, automatic consequences, and underscores the pervasiveness of mimicry as a response to experiences of ostracism (e.g., Zadro et al., 2004; Baumeistier & Leary, 1995; Macdonald and Leary, 2005).

Moreover, the result that the degree of post-ostracism increased lexical entrainment was similar when participants interacted with a partner who ostracised them versus a new innocent partner suggests that ostracism-induced affiliation is targeted to increasing social affiliation in general, and not only to repairing a particular social relationship. Importantly, this result supports previous evidence that post-ostracism effects are not partner-specific (e.g., Weerdmeester & Lange, 2019), which in turn strengthens the case for linguistic and non-linguistic mimicry being supported by the same mechanisms.

Furthermore, our results have implications for whether ostracism is a strong situation, i.e., whether its psychological and emotional consequences vary across individuals (see McDonald & Donnellan, 2012). One could argue that Experiment 1's null effect of neuroticism on lexical entrainment resulted from not using a large enough number of trials in our lexical entrainment task. Taking into account that participants answered several tasks in our ostracism experiments, we used a 15-trials version of the matching-and-naming task. Considering the small size effect of neuroticism on lexical entrainment in Experiment 1, and that we have previously shown that the test-retest reliability of our task increased as a function of its number of trials (see Chapter 3: Follow-up Analyses), it is possible that a 15-trial version of this task could not capture a large enough degree of individual variation for our regression analysis to indicate a significant relationship between neuroticism and post-ostracism entrainment in Experiment 1. However, the degree of lexical entrainment varied to a wide extent across participants (0-100%), suggesting that the task indeed captured individual variation in the tendency to entrain (see also Tobar-Henríquez et al., 2019). Taken together with Experiment 2's marginal effects of neuroticism on post-ostracism lexical entrainment, and the significant effect shown by our combined analysis,

a more likely explanation is that Experiment 1's null result occurred due to a lack of statistical power.

Thus, our neuroticism results suggest that the effects of ostracism may indeed be mediated by individual differences in neuroticism, in turn suggesting that ostracism is not a strong situation that inflicts the same degree of emotional and psychological distress across individuals. But what might determine the extent to which ostracism effects are moderated by neuroticism? Consistent with previous research suggesting a link between neuroticism and increased needs for social approval (Eysenck and Eysenck, 1991), we suggest that this relationship may reflect an over-compensatory measure. In particular, experiencing ostracism may have inflicted such a high degree of emotional distress to high-neuroticism individuals versus low-neuroticism individuals, that it enhanced social affiliative dispositions as an over-compensatory measure to regain social belonging. Future research could further our understanding of the underlying components of the relationship between neuroticism and post-ostracism lexical entrainment by examining, for example, the extent to which the degree of post-ostracism lexical entraiment is predicted by self-reported ostracism-induced emotional distress or disatisfaction of basic psychological needs (e.g., Williams et al., 2002).

In addition, our results inform previous theories of behavioural mimicry. In particular, they suggest that both non-linguistic and linguistic mimicry may be elicited due to increased social affiliative goals, supporting that both kinds of mimicry share a social affiliation component. This finding indicates that, in principle, non-linguistic and linguistic mimicry may share other underlying mechanisms too. Future studies on lexical entrainment could illuminate this question by interrogating whether the degree of lexical entrainment might be predicted by degrees of non-linguistic behavioral mimicry (e.g., mimicry of facial expressions of emotions) and their underlying constructs (e.g., measures of social competence, e.g., Mauersberger et al., 2015). For example, are individuals who are more likely to lexically entrain also more likely to mimic a partner's facial expressions? If so, are both their tendencies to lexically entrain and to mimic their partner's emotional facial expression correlated with the same potential underlying mechanism (e.g., social competence)?

Critically, our results also speak to previous theories of lexical entrainment. First, they replicate previous findings that lexical entrainment occurs even when not interacting face-to-face with a partner (e.g., Brennan, 1991, 1996; Branigan et al., 2011; Tobar-Henríquez et al., 2019), supporting the experimental reliability of lexical entrainment and suggesting that the underlying mechanisms of lexical entrainment are deployed even in communicative situations with low degrees of interaction.

More importantly, our results provide novel evidence for a causal relationship between increased social affiliation and increased liklihood of lexical entrainment. In particular, in the light of previous evidence that ostracism increases affiliative behaviours in general (Carter-Sowell et al., 2008) and non-linguistic affiliative mimicry in particular (Lakin et al., 2008), the fact that ostracism increased the likelihood of entraining suggests that the tendency to entrain varies as a function of social affiliation goals. In turn, this finding supports that speakers may vary their lexical choices based on their emotional states and their need for social belonging, suggesting that language processing interacts with emotional processing during language use.

The positive relationship between ostracism and lexical entrainment also speaks to previous research showing that lexical entrainment reflects individual differences in language processing (Tobar-Henríquez et al., 2019). The fact that ostracism affected lexical entrainment at the group-level, taken together with the finding that post-ostracism lexical entrainment was predicted by individual differences in neuroticism, suggests that, in principle, individual differences in lexical entrainment may correlate with social affiliation skills and dispositions. Future research may address this by interrogating, for example, if lexical entrainment predicts individuals' attunement to social affiliation in non-linguistic settings (e.g., Carter-Sowell et al., 2008; Maner et al., 2007; Williams & Sommer, 1997).

In sum, we found that experiencing ostracism increased speakers' tendency to reuse their partner's words, suggesting that the scope of the effects of being ostracised are not restricted to non-functional behavioural mimicry only, but it permeates functional behaviours too, such as language use during dialogue. Importantly, our neuroticism results suggest that ostracism is not always a strong situation; its emotional effects and compensatory responses may vary across individuals, depending on

their basal social affiliative needs. Critically for theories of lexical entrainment, ostracism effects on lexical entraiment imply a social affiliation component to the tendency to reuse a partner's words and suggest that individual differences in lexical entrainment may be at least partially explained by basal social affiliation dispositions. This in turn suggests that speakers may vary their lexical choices depending on their emotional states and social goals, and that lexical processing interacts with social and emotional information during language use, opening new research directions to understand which mechanisms restrict and inform speakers' referential expressions.

Chapter 6

# 6. Speakers extrapolate community-level knowledge from individual linguistic encounters

## 6.1. Introduction

Language use is fundamentally variable: The same object can be called a *potato* or *spud* in English, or a *patata* or *papa* in Spanish. This variation reflects a combination of both individual-level and community-level influences. At the individual level, for instance, speakers' referential choices are strongly influenced by their personal history with their interlocutor (*personal common ground;* Clark, 1996). For example, speakers are more likely to use *papa* for a potato if their conversational partner previously used that name, a phenomenon known as *lexical entrainment* (Brennan & Clark, 1996; Branigan, Pickering, Pearson, McLean, & Brown, 2011).

But language users are not just individuals acting in isolation – they form part of larger speech communities with shared patterns of language use (e.g., Labov, 1972). For example, Castilians not only know that a potato can be referred to as *patata* or *papa*, but are also aware of their own speech community's preference for *patata* over *papa* (*communal common ground*; Clark, 1996). However, speakers' knowledge of other speech communities (e.g., the favored term for potato in Latin America) is likely to be more fragmented. Investigating how speakers use language with conversational partners from other speech communities offers an opportunity to better understand how interpersonal- and community-level influences interact, and moreover how people extrapolate from interpersonal-level experiences of language use to establish community-level knowledge of language preferences.

In this paper, we investigate how speakers adapt their referential expressions based on interpersonal- and community-level factors. First, we examine how speakers' referential expressions are shaped by their partner's previous usage during an interaction, and whether beliefs about that partner's speech community modulate these effects. Second, we examine how speakers extrapolate community-level lexical knowledge from these individual encounters, and the role of their partners' community membership in constraining these extrapolations.

Lexical entrainment is a key phenomenon for investigating the variability of referential expressions. Such entrainment may arise in part as a result of recent linguistic processing, so that a partner's use of *papa* makes the term accessible in a speaker's memory, thus enhancing its retrieval and reuse (Pickering & Garrod, 2004; Horton & Gerrig, 2005, 2016; Neely, 1976; Meyer, 1996). But it may also reflect a conceptual pact between interlocutors, that is a belief that they share the same perspective on the referent (audience design). Under this logic, speakers reuse a partner's term to accommodate their language use to their partner's expectations, thus facilitating mutual comprehension (Brennan & Clark, 1996; Clark, 1996; Branigan et al., 2011). Importantly, audience design accounts do not necessarily imply that speakers make conscious rational inferences about their partner's knowledge; this process could as well be automatic and/or unconscious (e.g., Brown-Schmidt, 2009; Brown-Schmidt, Yoon, & Ryskin, 2015; Horton & Gerrig, 2005, 2016).

Although personal common ground has been the focus of most audience design research, speakers can also rely on linguistic communal common ground when engaging in audience design. Here, we focus on linguistic communal common ground shared by geographically-defined speech communities, e.g., defined by continent (Castilian Spanish versus Latin American Spanish), by country (Castilian variety, Mexican variety, Argentinian variety), and so on.

Clearly, speakers must build up knowledge of their community's communal ground (i.e., *in-community knowledge*) through individual encounters with other members from their community (i.e., *in-community partner*), gradually learning which language usages are commonplace and which idiosyncratic. For instance,

through their interactions with in-community (Castilian) partners, Castilian children will learn Castilian community preferences, developing in-community-level knowledge; e.g., that *patata* is the Castilian favored term for potato, though some individual Castilian speakers may prefer *papa* (Clark, 1996; Clark, 2007). Interestingly, previous work suggests that the more input we have had from different partners, the less sensitive our linguistic knowledge is to new input (e.g., Lev-Ari, 2018). This in turn suggests that possessing solid in-community-level knowledge about their own community preferences, based on numerous linguistic encounters with in-community partners, enables speakers to distinguish idiosyncratic in-community partners' preferences from in-community-level preferences.

These considerations may constrain lexical entrainment. If an in-community speaker names an object with a disfavored term (e.g., *papa*) instead of a favored term (i.e., *patata*) when interacting with an in-community addressee, then that addressee will know that this usage is that particular speaker's idiosyncratic preference, which does not match their own community-level preference. But all members in the same speech community would be expected to be familiar with (hence, understand) their own community's favored terms. Hence hearing a Castilian speaker use this disfavored term (*papa*) would not necessarily motivate a Castilian partner to reuse that term in order to enhance communication. Moreover, experiencing an in-community speaker using a disfavored term would not meaningfully update the partner's in-community knowledge, given their substantial prior experience with other in-community speakers – and so would leave unchanged their likelihood of using that disfavored term with a subsequent in-community speaker (aside from any transient low-level priming effects that might promote its use).

Similarly, through individual linguistic interactions with members of other communities such as Mexicans (i.e., *out-community partners*), Castilians can establish knowledge about those communities, for example that *papa* (the disfavored term in their own community) is in fact the favored term for potato in Mexico. Almost always, speakers' knowledge about other communities' preferences (i.e., *out-community knowledge*) will be rooted in fewer linguistic encounters than their knowledge about their own community-level preferences, and will thus be less

accurate (Clark, 1996; Clark, 2007; Lev-Ari, 2018). Consequently, it will be harder for an individual to discern whether an out-community partner's term is an individual or a community preference. For example, if a Mexican used a disfavored term (e.g., *papa*), a Castilian's knowledge of Mexican community preferences might not be accurate enough for them to discern whether *papa* is an idiosyncratic use or an out-community-level (Mexican) preference. Thus, an out-community partner's use of a disfavored term may increase speakers' disposition to entrain, in order to enhance communication with that particular out-community conversational partner.

Importantly, given their limited experience with out-community partners' community-level preferences, speakers may be more sensitive to new input (Lev-Ari, 2018), so that a single linguistic interaction with a Mexican may update a Castilian's knowledge of Mexicans' community-level preferences, thus strengthening the belief that Mexicans have a community preference for the Castilian disfavored term (*papa*) over the favored term (*patata*). This will in turn increase the Castilian's likelihood to use that disfavored term (*papa*) with subsequent Mexican partners, as a result of the association of the term with out-community partners' assumed preferences (alongside any transient low-level priming effects). However, this prediction has yet to be tested.

Consistent with accounts that emphasise the role of common ground, lexical entrainment is influenced by speakers' beliefs about a partner's community membership. In particular, Branigan et al. (2011) had speakers of British English complete a computerised matching-and-naming task where critical items were pictures of objects that could be named with a favored term (e.g., *potato*) and a disfavored (but still acceptable) term (e.g., *spud*) in British English (established via a pretest). The confederate was in reality pre-programmed software, which always named critical items before participants; lexical entrainment was then measured as the proportion of trials on which participants used the same disfavored term as the partner had used before. Critically, Branigan et al. manipulated participants' beliefs about whether their 'partner' was a computer or a human, and if the partner was a computer, whether it was more or less capable. They found not only that participants entrained more often to computer-partners than to human-partners, but also that they

entrained more often to ‘less capable’ computer-partners than to ‘more capable’ computer-partners, suggesting that participants entrained more when they were less confident about their partner’s understanding of the favored term.

These results are congruent with the hypothesis that the less experience we have of a speech community’s preferences, the more likely we are to make inferences about a partner’s community preferences from individual linguistic encounters with members of that partner’s community. People typically have considerably less linguistic experience with computers than with people, and therefore have weaker models of community preferences for computers than for humans (Branigan, Pickering, Pearson & McLean, 2010). Therefore, participants could potentially interpret an individual computer’s use of a disfavored term as representative of a computer-community preference, and then assume that computers might not understand the favored term, increasing the likelihood of lexical entrainment on the disfavored term. In contrast, participants’ extensive previous experience with native English-speaking humans would mean that they would interpret a human’s use of a disfavored term as an idiosyncratic preference, and would not assume that their human-partner would not understand the favored term.

So far, we have focused on the extent to which speakers reuse a partner’s term while interacting with the same partner. But people tend to reuse previously used terms (or *maintain precedents*) even when interacting with a new partner. Brennan and Clark (1996)’s Experiment 3 had participants play two consecutive sessions of an interactive referential task (see also Krauss & Glucksberg, 1969; Clark & Wilkes-Gibbs, 1986), and measured the proportion of second session trials where participants maintained a term that they had used in the first session. Importantly, they manipulated whether the first session terms were part of linguistic personal common ground during the second session: Participants either interacted with the same partner throughout the task or swapped partner between sessions. Participants tended to maintain first session terms in the second session even when interacting with a new partner, consistent with facilitation due to transient priming effects.

Critically, however, participants were more likely to maintain their first session terms when these terms were part of personal linguistic common ground than

when they were not, i.e., when they interacted with the same partner in both sessions than when they switched partners. This finding of partner-specific referential maintenance has been repeatedly replicated (e.g., Horton & Gerrig, 2002, 2005), and is supported by research showing that comprehenders experience more difficulty understanding a partner's new term for a referent when that partner has previously used another term for it than when they have not (e.g., Metzing & Brennan, 2003; see Krönmuller & Barr, 2015).

In sum, speakers' referential expressions are influenced by individual experiences with specific partners, in ways that reflect effects of both recent lexical processing and beliefs about partners' likely expectations of language use. However, it remains unknown how exactly beliefs about a partner inform and constrain speakers' referential expressions during dialogue. In particular, it is unclear what role beliefs about a partner's speech community might play in entrainment, and whether community information about a partner is encoded along with lexical information during dialogue. More specifically, do speakers encode partner's community information during individual linguistic experiences and extrapolate from these individual experiences to inferences about likely community preferences that might affect their referential expressions with future partners?

### 6.1.1. The present experiments

To address these issues, we investigated the effects of participants' beliefs about their partner's speech community on both lexical entrainment (i.e., speakers' initial tendency to reuse a partner's term) and maintenance of entrained terms (i.e., speakers' subsequent reuse of that term). In two internet-based experiments, native speakers of (varieties of) Spanish engaged in two sessions of an interactive online picture matching-and-naming task, based on Branigan et al. (2011). Spanish provides an excellent test-case, because of its many different regional varieties. Participants took turns with a partner to either select a picture named by their partner or name a picture for their partner to select. Critically, the 'partner' was pre-programmed software and participants believed that they were interacting with different partners in each session. Experimental trials comprised a picture of an object (e.g., a potato)

that could be named in participants' speech community with both a favored term (e.g., *patata)* and a disfavored but acceptable term (e.g., *papa*) (as established by a pretest).

In the first session, the partner always named the target first, using the disfavored term, and the participant named the same target in a subsequent turn. We measured lexical entrainment as the proportion of trials where participants reused the same disfavored term. In the second session, participants named the same targets for their new partner (but their partner never named them). We measured maintenance of entrained terms as the proportion of trials where participants maintained a disfavored name that they had used in the first session. Critically, in the first session we manipulated participants' beliefs about whether their partner was an in-community partner (i.e., from their own speech community) or an out-community partner (i.e., from another speech community); in the second session, we manipulated their beliefs about whether their second partner was from the same community as their previous partner or not.

We examined lexical entrainment in the first session, and maintenance of entrained terms in the second session. Based on previous research, we expected that participants would lexically entrain with their partner, such that they would use the disfavored term to refer to an object more often after their partner had used it than in the spontaneous naming task used to norm our materials. But we expected that the extent to which participants lexically entrained would differ between conditions. If lexical entrainment is influenced by beliefs about a partner's community membership, such that speakers are more likely to entrain on a disfavored term when they are less confident about their partner's community preferences (and conversely less likely to entrain when they are more confident about their partner's community preferences), then participants should entrain more often to out-community partners than to in-community partners.

We similarly expected that participants would tend to maintain entrained terms, so that they would be more likely to use the disfavored term after they had previously used it than in a spontaneous naming task. However, we predicted that, if speakers encode their partner's community information during lexical processing,

their tendency to maintain entrained terms would vary between conditions based on their partners' community memberships. If maintenance of entrained terms reflects speakers' extrapolation of communities' preferences - such that speakers are more likely to maintain a (disfavored) entrained term when they are less confident about their partner's community preferences (and conversely less likely to maintain it when more confident about their partner's community preferences) - then participants should maintain an out-community partner's term more often when subsequently interacting with another out-community partner from the same community than with an in-community partner (i.e., from their own speech community). However, they should not maintain an in-community partner's term more often with another in-community partner than with an out-community partner.

## 6.2. Experiment 1: Do speakers extrapolate community lexical preferences from individual linguistic encounters?

Experiment 1 used a two-session picture-matching-and-naming task to investigate the effects of beliefs about a partner's community membership on both lexical entrainment to a first partner, and maintenance of entrained terms from a first to a second partner. We measured whether Castilian participants reused a disfavored term that their partner had used earlier (Session 1; Lexical Entrainment), and whether they reused disfavored terms that they had previously used in Session 1 when interacting with a new partner (Session 2; Maintenance of Entrained Terms). In Session 1, we manipulated participants' beliefs about whether their partner was from their own community or another community (First Partner's Community: *In-Community Partner* [Spain] vs. *Out-Community Partner* [Latin America]). In Session 2, we manipulated participants' beliefs about whether the second partner was from the same community as the first partner or not (Second Partner's Community*: Same as First Partner* vs. *Different from First Partner*).

If Lexical Entrainment is affected by beliefs about their partner's speech community, participants should entrain more with out-community partners than in-community partners. If Maintenance of Entrained Terms reflects learning of

community-appropriate terms, participants should generalize (i.e., maintain) disfavored terms from an out-community partner to another out-community partner from the same community more often than to an in-community partner, but they should not generalize (i.e., maintain) disfavored terms from an in-community partner to another in-community partner more often than to an out-community partner.

### 6.2.1. Method

**Participants.** We recruited 160 online participants through the portal Prolific [https://prolific.ac/]. To be included, participants had to be native speakers of Castilian Spanish, born and raised in Spain, and aged between 18 and 50 years old (M=32, SD=8). Participants were paid £2. Ethical approval for the experiments reported below was obtained from the Psychology Research Ethics Committee of the University of Edinburgh (429-1718).

**Design.** We used a 2 (First Partner's Community: *In-Community Partner* vs. *Out-Community Partner*) x 2 (Second Partner's Community: *Same as First Partner* vs. *Different from First Partner*) between-participants and within-items factorial design. The dependent variables were (a) Lexical Entrainment (i.e., participants' use of disfavored terms in Session 1) and (b) Maintenance of Entrained Terms (i.e., participants' reuse during Session 2 of disfavored terms that they had used in Session 1).

**Materials.** In both sessions, we used 21 target pictures that could be labelled with either a favored term or a disfavored but acceptable term in Castilian Spanish , as well as 21 unambiguous filler pictures.

To create the experimental items, we conducted two norming tasks, each involving a different set of participants drawn from the same population as those in the main experiment. Participants were paid £1. A separate group of 110 native speakers of Castilian Spanish (aged 18-60, M=29, SD=7) answered two questions in an online survey (via Prolific). For each of 119 pictured objects, the first question elicited a favored term (What is the first word you would use to name this object? [*¿Cuál es la primera palabra que se te ocurre para nombrar este objeto?*]), and the

second question elicited a disfavored term (If you had to use another word, which one would you use? [*Si tuvieras que usar otra palabra, ¿cuál usarías?*]).

From these ratings, we gathered 60 potential target pictures, for which at least 70% of participants had provided the same favored term, and at least 10% of participants had provided the same disfavored term. The 60 potential targets were then entered into a second rating task, in which 60 new Castilian individuals (aged 18-50, M=33, SD=9) rated the acceptability of these disfavored terms with respect to the pictures, on a scale from 1 to 7. We used this to create the final set of 21 disfavored terms (see Appendix), each of which had been spontaneously used with a frequency below 30% when answering the favored term question (M=4%, SD=6%) and had an acceptability rating above 5.3 in the acceptability rating task (M=6.1, SD=0.5). We also used the favored term task to choose 21 filler pictures, in which 80% of participants agreed on the same favored term.

**Procedure.** Participants completed two sessions of an online matching-and-naming task. On each trial, they were shown two pictures (see Figure 6.1), and they then either clicked on the target picture named by their partner (matching trials) or typed the term of the indicated target picture (naming trials). In matching trials, participants were given feedback on their response: They saw either a ‘well done’ or ‘wrong answer’ message when they matched the right (target) picture or the wrong (distractor) picture, respectively. In naming trials, they received feedback on their partner’s matching choice, which was always positive. In Session 1, half of the trials were filler trials, on which the target picture only had a single name (e.g., *limón* [lemon]). The other half were experimental trials, on which the target picture could be named with either a favored term (e.g., *patata* [potato]) or a disfavored, but still acceptable, term (e.g., *papa*).

The structure of the matching and naming task is illustrated in Figure 6.1B. In Session 1, participants alternated matching and naming trials with a ‘remote player’, in fact pre-programmed software that provided scripted answers. Importantly, the trial structure in Session 1 meant that the software always named the experimental targets before the participants, using the disfavored terms (e.g., *patata*). In Session 2, only participants named the experimental targets (presented in the same order as in

Session 1); importantly, in Session 2, the 'partner' never named experimental targets (and instead named only fillers).

Critically, to manipulate participants' beliefs about their partner's speech community membership, we explicitly told participants whether they would play with an in-community partner (i.e., a Castilian partner) or out-community partner (i.e., a Latin American partner) in each session. They were told both partners' speech community membership before the task began, and were reminded of each partner's membership at the beginning of each session. Importantly, we ran a manipulation check at the end of the second session and excluded data from participants who reported playing with partners from a community different to the ones they had been told in each condition.

Participants were recruited to take part in this study on Prolific, using an advertisement that was visible only to individuals who met our inclusion criteria (see above). The advertisement stated that participants would play two sessions of a picture matching-and-naming task, and that they would play with a different remote player in each session. We also made explicit the speech community of each of the two partners with whom participants would play (e.g., You will play with two different partners from {Spain/Latin America} [*Jugarás con dos jugadores remotos, ambos serán de España (América Latina)*]*;* You will play with a partner from {Spain/Latin America} in Session 1, and a partner from {Spain/Latin America} in Session 2 [*Jugarás con una persona de {España/América Latina} en la primera sesión y con una persona de {España/América Latina} en la segunda sesión*]).

Participants then completed a consent form, and were told to wait to be matched with a remote player; at this point they were reminded of the first partner's linguistic community (e.g., We are connecting you to a partner from {Spain/Latin America}. Thank you for your patience [*Te estamos conectando con un jugador de {España/América Latina}. Muchas gracias por tu paciencia*]). After two minutes, they were redirected to the first task (programmed with JSPsych and available at https://github.com/anitatobar/lingcommunities; de Leeuw, 2015), where they were asked to alternate turns with their partner to match and name one out of two pictures that would appear on the screen. In each trial, they saw two pictures and were asked

to either wait for their partner's response so that they could select the correct (matching) picture, or to name the picture to the right or left (depending on where the target appeared, which was randomised) (see Figure 6.1).

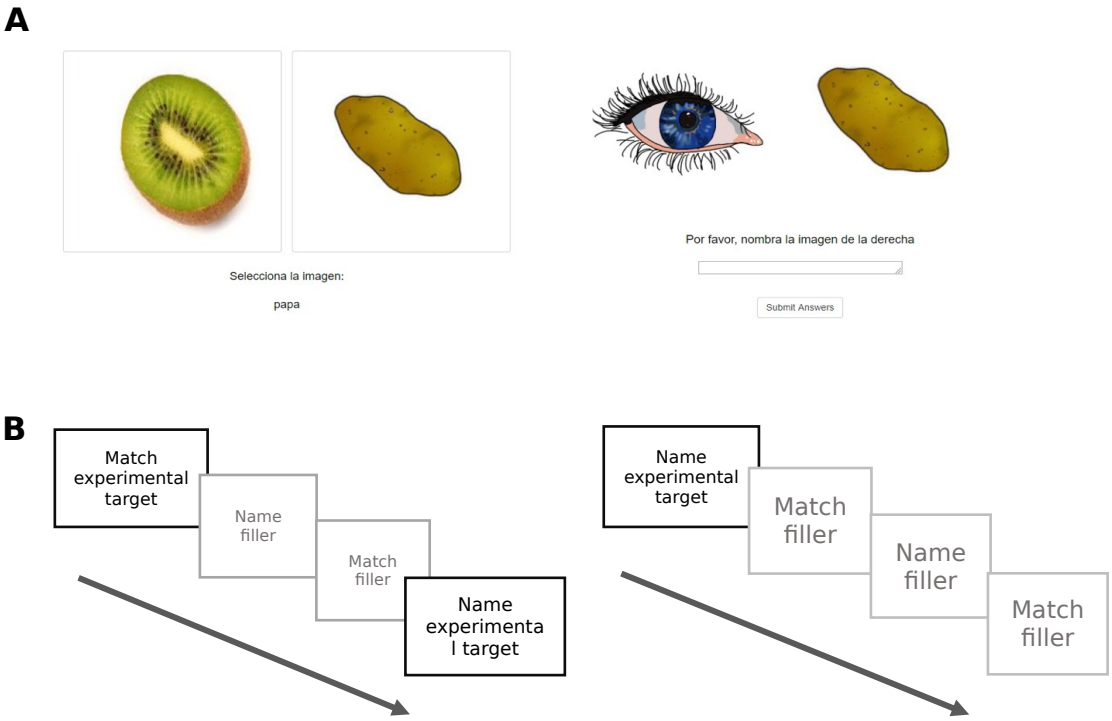


**Figure 6.1. A**. Examples of matching and naming trials (where the favored term is *patata* and disfavored term is *papa*). In matching trials (left), the participant selected the named picture. In naming trials (right), they named the target picture, which was presented along with a randomly selected distractor. **B.** Sequence of experimental items and fillers presentation. In Session 1 (left), participants first matched an experimental target picture with the corresponding disfavored term; they subsequently named and matched two fillers; and finally named the previously matched experimental target. In Session 2 (right), participants first named an experimental target (already named in Session 1) and then encountered three fillers; they never experienced their partner naming the experimental target. Participants played with two different partners across the two sessions.

After matching and naming the 42 items, they were told to wait to be matched to a new remote player and were again reminded of the partner's linguistic community (e.g., We are connecting you to a new partner from {Spain/Latin America}. Thank you for your patience [*Te estamos conectando con un nuevo jugador de {España/América Latina}. Muchas gracias por tu paciencia*]). After two minutes, they were told the new partner from either Spain or Latin America was waiting for them and were asked to press a key to start the task (i.e., A new partner from {Spain/Latin America} is waiting for you. Press any key to start the game [*Tu compañera/o de {España/América Latina} te está esperando para comenzar el juego. Aprieta cualquier tecla para comenzar*]). During the second session, participants encountered the same experimental targets they had encountered in Session 1, presented along with randomised distractors, and interspersed with fillers presented in a randomised order. Importantly, participants did not experience their partner naming the target during Session 2.

After the task, participants were redirected to a survey, where we checked participants' beliefs about their partner by asking 'How many people did you play with and where were they from?' [*¿Con cuántas personas jugaste y de dónde eran?*]; we coded whether participants reported playing with multiple partners, where the partners were from, and whether participants explicitly indicated that they suspected they had played with a computer. Finally, participants were redirected to a Prolific website and received a completion code in order for us to confirm their payment.

### 6.2.2. Results

**Data processing and exclusions.** In Session 1, we coded all naming trials for whether they showed Lexical Entrainment (using the disfavored term used by the partner) or not (using any other Castilian term to name the target). In Session 2, we coded all naming trials for whether they showed Maintenance of Entrained Terms (maintaining an entrained term that the participant had previously used in Session 1) or not (using any other term). Occasionally (less than 10% of trials), participants named or selected the distractor instead of the target; these trials were coded as NA.

No participants reported believing that they had not played with a real person or that they had played with the same partner in both sessions. Critically, no participants reported believing that they had played with partners from a speech community different to the one they had been told in each condition.

**Analyses.** We conducted separate analyses for Lexical Entrainment and Maintenance of Entrained Terms. All analyses were carried out in the R programming language and environment (R Development Core Team, 2016). We tested the effects of independent variables using mixed-effect logistic regressions, using lme4 package version 1.1 (Bates, Maechler, Bolker, Walker, Christensen, Singmann, Dai, Grothendieck, & Green, 2015)[5]. All dependent factors included as fixed effects were sum coded (i.e., -1, 1), and we always used the maximal random structure justified by our design that allowed the models to reach convergence (Barr, Levy, Scheepers, & Tily, 2013). To assess the significance of all main effects and interactions involving fixed factors, we used Wald tests. We report results for key regression coefficients in the main text and full regression model results in tables; full model structures are also reported in the table captions. Moreover, for key null results we report Bayes Factors, which quantify the likelihood of observing a given dataset if there were no difference across conditions, compared to if there were a difference (Wagenmakers, 2007). The analysis scripts including all models (even those that did not reach convergence) are available at https://github.com/anitatobar/lingcommunities.

We conducted two analyses on participants' Lexical Entrainment in Session 1. First, we assessed the overall presence of an entrainment effect using a Wilcoxon test; we tested whether the produced proportion of disfavored terms was higher during Session 1 compared to during the spontaneous naming task used to norm the materials. Then, we tested whether participants' beliefs about their partner's speech community affected entrainment, by regressing the use of disfavored terms in

[5] Tables show models' structure using R syntax. From left to right, the first argument corresponds to the dependent variable, the second argument represents fixed effects, and arguments in parentheses are variables added as random effects (in our studies, by participants and items). In particular, random intercepts are represented with the number 1, and random slopes are represented by variables added to the right to the intercept, e.g., (1 + *slope* | items).

Session 1 against First Partner's Community (i.e., *In-Community Partner* versus *Out-Community Partner*).

We conducted two analyses on participants' Maintenance of Entrained Terms in Session 2. First, we assessed whether participants maintained referential precedents, using two Wilcoxon tests: We compared the proportion of disfavored terms produced in Session 2's naming trials both against the norming task and against Session 1's naming trials. The second analysis focused only on Session 2 trials where participants had previously used disfavored terms in Session 1 (i.e., lexical entrainment trials). In particular, we tested whether maintenance of disfavored terms was affected by participants' beliefs about their partner's speech community, by regressing the use of disfavored terms on lexical entrainment trials against the interaction between First Partner's Community (*In-Community Partner* versus *Out-Community Partner*) and Second Partner's Community (*Same as First Partner* versus *Different from First Partner*).

**Lexical Entrainment.** In Session 1, participants used the disfavored term on approximately half of the trials (50%[SD=30%] by-participants and 52%[17%] by-items). The by-items Wilcoxon test indicated that disfavored terms were used significantly more frequently in Session 1 than during the spontaneous naming task used to norm the materials (4%[6%]; $V=0$, $p<.0001$), clearly suggesting a Lexical Entrainment Effect (see Figure 6.2).

Critically, the degree to which participants lexically entrained was not significantly affected by their First Partner's Community (see Table 6.1, $\beta = .038$, $SE=.15$, $z=.25$, $p>.05$). Participants used entrained terms at similar rates whether they believed their partner was from another speech community (52%[29%]) or from their own speech community (49%[32%]), suggesting that lexical entrainment might not be affected by a partner's speech community during remote, computerised referential tasks (see Figure 6.2).

To confirm that our data supported the null hypothesis of no difference in participants' tendency to entrain to partners from their own community versus to partners from another community, we calculated a Bayes Factor over a model assuming no difference between conditions (null model) and a assuming a difference between

conditions (alternative model). The null model included a fixed intercept, and random intercepts by items and participants. The alternative model included First Partner's Community as main effect, and random intercepts by items and participants. We used the two models' Bayesian Information Criterion (BIC) values to estimate the Bayes Factor as e(BIC_alternative – BIC_null)/2 (see Wagenmakers, 2007, and Masson, 2011). The null model fit the data slightly better, by a Bayes Factor of e(3571.64-3563.603)/2 = 55.63, with a posterior probability in favour of the null model (BF / (BF + 1) = .98), which provides strong evidence against the hypothesis that speakers entrain more to out-community partners than in-community partners (Raftery, 1995).

We also tested whether participants' tendency to entrain was affected by the Second Partner's Community (note that they had not yet interacted with the second partner), to check for baseline differences across the four conditions (Table 6.1). We found no evidence for such differences: Participants entrained to their first partner at similar rates when their second partner was going to be from that partner's community (48%[31%]) or from a different community (51%[28%]), and there was no significant interaction between First Partner's Community and Second Partner's Community.

*Table 6.1. Experiment 1 - Session 1 Trials: LexicalEntrainment ~ FirstPartner + SecondPartner + FirstPartner:SecondPartner + (1|participants) + (1 |items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | 0.08557 | 0.22206 | 0.385 | 0.700 |
| First Partner's Community | 0.04121 | 0.14635 | 0.282 | 0.778 |
| Second Partner's Community | 0.03382 | 0.14636 | 0.231 | 0.817 |
| First:Second | -0.10009 | 0.14634 | -0.684 | 0.494 |

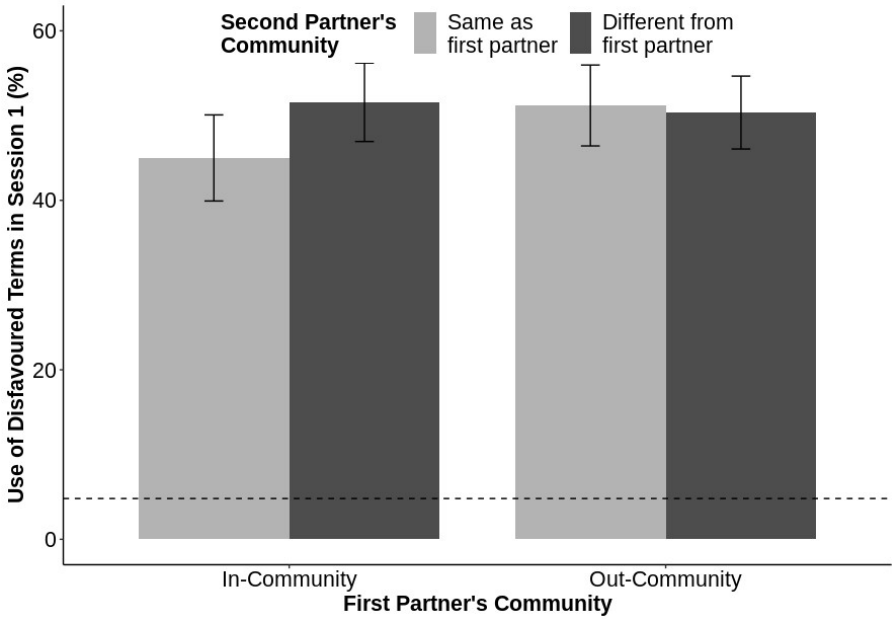


**Figure 6.2.** Mean and standard error of the percentage of use of disfavored terms in Session 1 (y-axis) across First Partner's Community (x-axis) and Second Partner's Community (color-coded). The horizontal dashed line represents the mean of percentage of use of disfavored terms on the pretest.

**Maintenance of Entrained Terms.** Overall, participants reused disfavored terms on approximately one third of trials (34%[27%] by-participants, and 33%[12%] by-items). A by-items Wilcoxon test indicated that participants maintained entrained terms: They were more likely to reuse (entrained) disfavored terms in Session 2 than participants who named the objects spontaneously during the norming task (4%[6%]; V=0, p<.0001; see Figure 6.3). However, they did not maintain terms perfectly, so that participants were significantly less likely to reuse disfavored terms in Session 2 than to initially use such terms in Session 1 (V=0, p<.0001).

For items on which the disfavored term was used in Session 1, participants maintained that disfavored term on 70%[27%] of trials in Session 2. Figure 6.3 illustrates how this rate varied depending on the speech communities of the two partners. There was no significant effect of Second Partner's Community (see Table 6.2): Participants maintained disfavored terms at similar rates across two partners from differ-

ent communities (68%[29%]) as across two partners from the same community (72%[26%]).

There was a significant effect of First Partner's Community (β=-.36, SE=.13, z=-2.7, p<.01, see Table 6.2): Participants maintained disfavored terms more often in Session 2 when their Session 1 partner was an in-community partner (76%[25%]) than an out-community partner (64%[28%]). But critically, these effects were qualified by a significant interaction between First Partner's Community and Second Partner's Community (β=-.3, SE=.13, z=-2.3, p=.02, see Table 6.2 and Figure 6.3). That is, participants' tendency to maintain an entrained term with a second partner was modulated by the second partner's community, relative to the first partner's community.

*Table 6.2. Experiment 1- Session 2 Trials: Maintenance ~ FirstPartner + SecondPartner + FirstPartner:SecondPartner + (1|participants) + (1|items)*

| Predictors | Estimate (β) | Std. Error (SE) | z-value | P-value |
|---|---|---|---|---|
| (Intercept) | 0.88371 | 0.18554 | 4.763 | **<.001** |
| First Partner's Community (First) | -0.35707 | 0.13115 | -2.723 | **0.00647** |
| Second Partner's Community (Second) | -0.07909 | 0.13096 | -0.604 | 0.54586 |
| First:Second | -0.30301 | 0.13106 | -2.312 | **0.02078** |

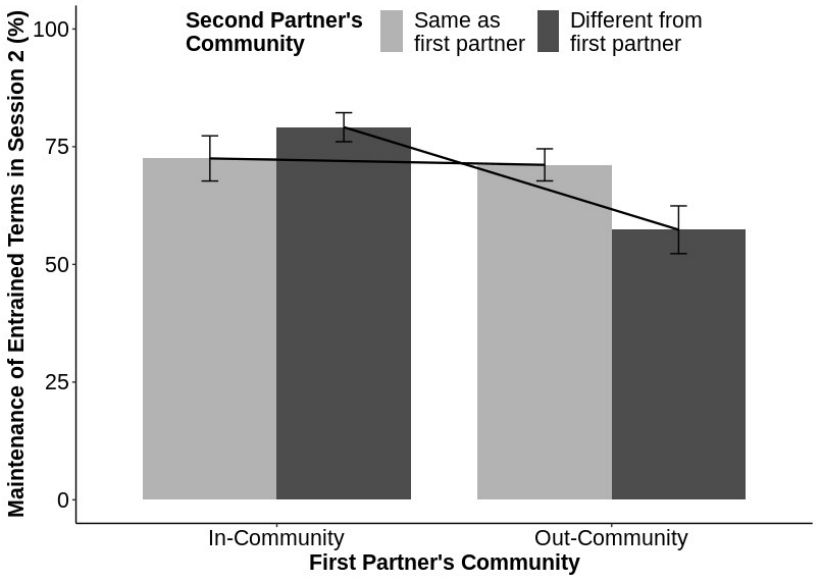


**Figure 6.3.** Mean and standard error of percentage of (re)use of disfavored terms in Session 2 (y-axis) across First Partner's Community (x-axis) and Second Partner's Community (color-coded). Solid lines represent a significant interaction between the First Partner's Community and Second Partner's Community on maintenance of entrained terms.

To understand this interaction, we subset our data and tested the effect of Second Partner's Community on trials on which participants had entrained to an in-community partner (in-community partner trials) versus an out-community partner (out-community partner trials). For in-community partner trials, there was not a significant effect of Second Partner's Community: Participants who first entrained to an in-community partner subsequently maintained those terms to the same extent with an in-community partner as an out-community partner (72%[30%] vs. 79%[20%]; $\beta$=0.21, SE=.19, z=1.1, p>.05, see Table 6.3 and Figure 6.3). In contrast, we found a significant effect of Second Partner's Community on out-community partner trials: Participants who first entrained to an out-community partner in Session 1 subsequently maintained those entrained terms less often with an in-community partner in Session 2 (57%[SD=32%]) than with an out-community partner (71%[21%]; $\beta$=-.38, SE=.18, z=-2.2, p=.027, see Table 6.4 and Figure 6.3).

*Table 6.3. Experiment 1 - In-community Partner Trials: Maintenance ~ SecondPartnersCommunity + (1|participants) + (1|items)*

| Predictors | Estimate (β) | Std. Error (SE) | z-value | P-value |
|---|---|---|---|---|
| (Intercept) | 1.2880 | 0.2335 | 5.515 | **<.0001** |
| Second Partner's Community | 0.2130 | 0.1922 | 1.108 | 0.268 |

*Table 6.4. Experiment 1 - Out-community Partner Trials: Maintenance ~ SecondPartnersCommunity + (1|participants) + (1|items)*

| Predictors | Estimate (β) | Std. Error (SE) | z-value | P-value |
|---|---|---|---|---|
| (Intercept) | 0.5423 | 0.2175 | 2.494 | **0.0126** |
| Second Partner's Community | -0.3878 | 0.1758 | -2.206 | **0.0274** |

Overall, these results suggest not only that speakers maintain entrained terms across partners, even during remote computer-mediated interactions, but also that this tendency is mediated by the speech communities of those partners. More specifically, participants extrapolated lexical knowledge from an individual linguistic interaction with an out-community partner to another out-community partner (from the same community) more often than to a partner from the speaker's own community; however, they extrapolated lexical knowledge from an in-community partner to other in-community partners at similar rates as to out-community partners.

Taken together, the results of Experiment 1 do not suggest that speakers' beliefs about a partner's speech community membership affect their likelihood of entraining on the disfavored term for an object – but they do suggest that such beliefs affect speakers' tendency to maintain an entrained term in a subsequent interaction with a new partner. This finding in turn suggests that speakers encode community information about their partner during lexical processing in ways that can affect their subsequent behavior even when it does not affect their concurrent behavior. Specifically, participants extrapolated referential expressions that had been used by out-com-

munity partners (and that they had themselves adopted) more often to subsequent partners from the same out-community than to partners from their own community; however, they did not extrapolate expressions that had been used by partners from their own community (and that they had themselves adopted) more often to subsequent out-community partners than to partners from their own community. The finding that participants extrapolated expressions in some circumstances suggests that they were sensitive to their partners' believed community during entrainment even if it did not affect entrainment itself, and the particular pattern of extrapolation suggests that knowledge of other communities' preferences is more strongly affected by new language input than knowledge of their own community's preferences.

In Experiment 2, we investigated the generalizability and replicability of these effects. It is possible that the effects and non-effects of community found here are particular to Castilian Spanish speakers. For example, due to historical factors, Latin American individuals might be expected to be familiar with Castilian lexical preferences (e.g., the preference for *patata* over *papa*). If so, this could have interfered with the effect of a partner's community membership on lexical entrainment: If Castilians believed Latin out-community partners to be familiar with Castilians' preferences, there would be no motivation for them to entrain more strongly to Latin out-community partners in order to enhance mutual understanding. Similarly, the same expectation might have led Castilians to extrapolate Castilian in-community partners' preferences to Latin out-community partners at similar rates as to other Castilian in-community partners.

Thus, in Experiment 2 we replicated Experiment 1, but in a Mexican population, examining their entrainment and maintenance of entrained terms to a different Latin speech community ( Argentinian Spanish) with whose lexical preferences they were unlikely to be familiar. If the effects of Experiment 1 were due to Castilians expecting (Latin American) out-community partners to be familiar with Castilians' favored terms, then we would expect to find both that Mexicans entrain more to (Argentinian) out-community partners than to (Mexican) in-community partners, and that they do not extrapolate (Mexican) in-community partners' expressions to (Argentinian) out-community partners to similar rates as to other (Mexican) in-com-

munity partners. If they instead reflected a general tendency for speakers to entrain irrespective of a partner's speech community, but to extrapolate preferences with regard to partners' speech community, then we would expect to find that Mexicans entrain to (Argentinian) out-community partners to the same extent as to (Mexican) in-community partners, and that they extrapolate (Mexican) in-community partners' expressions to (Argentinian) out-community partners at similar rates as to other (Mexican) in-community partners.

## 6.3. Experiment 2: Do the effects of partner's community membership on speakers' lexical expressions generalize across speech communities?

Experiment 2 was identical to Experiment 1, except that participants were native speakers of Mexican Spanish who believed that they were interacting with Mexican speakers or with Argentinian Spanish speakers.

### 6.3.1. Method

Unless detailed, Experiment 2 used the same design and procedure as Experiment 1.

**Participants.** We recruited 160 native speakers of Mexican Spanish, aged 18-50, born and raised in Mexico.

**Items.** We used the same norming tasks as in Experiment 1 to create experimental items normed for Mexican Spanish. 110 volunteer native speakers of Mexican Spanish (recruited via Facebook, aged 18-60, M=32, SD=7) provided favored and disfavored terms, and 100 new Mexican individuals (aged 18-54, M=31, SD=7) rated acceptability of the disfavored terms. We created a final set of 20 disfavored terms (see Appendix); mean use as a favored term was 4%[SD=7%] and

mean acceptability was 6.0[SD=0.6]. We also used the first rating task to choose 20 filler pictures, in which 80% of participants agreed on the same favored term.

**Procedure.** As in Experiment 1, participants completed two sessions of an online matching-and-naming task, where they took turns with a partner to match and name pictures of objects with a favored and a disfavored term. The structure of the matching and naming task was the same as in Experiment 1 (Figure 6.1B). In each session, we manipulated participants' beliefs about whether their partner was a member of their own speech community (i.e., Mexico) or from another speech community (i.e., Argentina).

### 6.3.2. Results

**Data processing and exclusions.** Coding and exclusions were carried out as in Experiment 1. Five participants reported believing that they had not played with a real person but no participants reported believing that they had played with multiple partners. Critically, no participants reported believing they had played with partners from a speech community different to the one they had been told.

**Lexical Entrainment.** As in Experiment 1, in Session 1 participants used the disfavored term on approximately half of the trials (54%[SD=31%] by-participants and 50%[17%] by-items). The by-items Wilcoxon test indicated again that the frequency of use of disfavored terms was significantly higher than in the norming task (4%[6%]; $V=0$, $p<.0001$), clearly suggesting the presence of a Lexical Entrainment Effect (see Figure 6.4).

As in Experiment 1, the degree to which participants lexically entrained was not significantly affected by First Partner's Community (see Table 6.5, $\beta = .15$, $SE=.16$, $z=.95$, $p>.05$). Participants were not more likely to entrain to a partner from another speech community (55%[32%]) than a partner from their own speech community (52%[30%]), again suggesting that lexical entrainment might not be affected by beliefs about a partner's community during this computerised referential task. Consistent with this, the Bayes Factor indicated that the null model fit the data slightly better, by a Bayes Factor of $e(3203 – 3210)/2 = 35.47$, and a posterior probability in favour of the null model $BF / (BF + 1) = .97$), which provides strong evid-

ence against the hypothesis that speakers entrain more often to out-community partners than in-community partners.

Importantly, a control analysis again found no significant differences in entrainment across the four conditions of the study (Table 6.5). Participants entrained to their first partner at similar rates when their second partner would be from that partner's community (53%[31%]) or a different community (54%[30%]), and there was no significant interaction between First Partner's Community and Second Partner's Community.

*Table 6.5. Experiment 2 - Session 1 Trials: LexicalEntrainment ~ FirstPartner + SecondPartner + FirstPartner:SecondPartner+ (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | 0.319716 | 0.231504 | 1.381 | 0.167 |
| First Partner's Community | 0.151154 | 0.159517 | 0.948 | 0.343 |
| Second Partner's Community | -0.004713 | 0.159534 | -0.030 | 0.976 |
| First:Second | -0.182111 | 0.159516 | -1.142 | 0.254 |

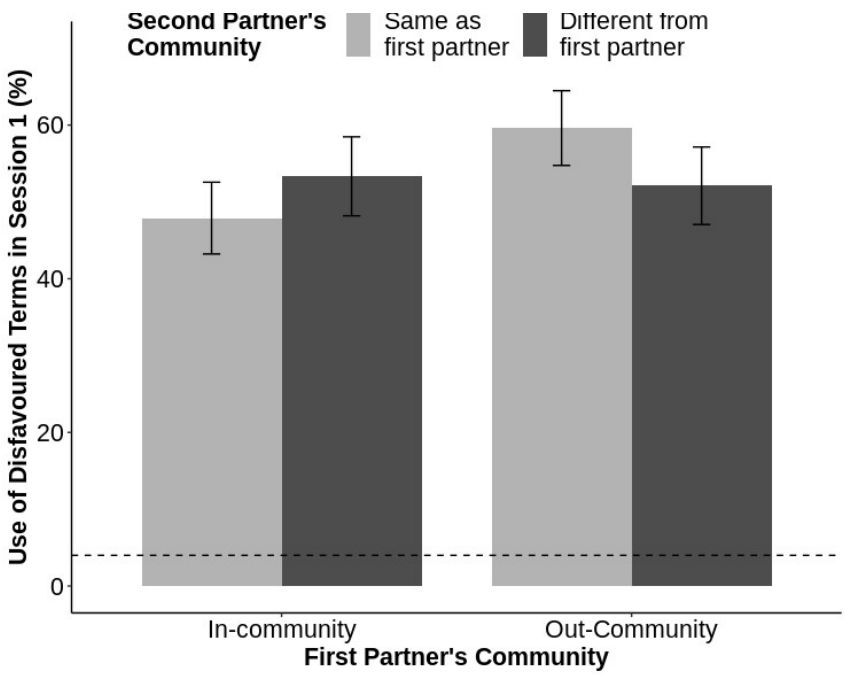


**Figure 6.4.** Mean and standard error of the percentage of use of disfavored terms in Session 1 (y-axis) across First Partner's Community (x-axis) and Second Partner's Community (color-coded). The horizontal dashed line represents the mean of percentage of use of disfavored terms on the pretest.

**Maintenance of Entrained Terms.** As in Experiment 1, in Session 2 participants used disfavored terms on approximately one third of trials (35%[26%] by participants and 33%[13%] by items). Again, the by-items Wilcoxon test indicated that participants maintained entrained terms: They were more likely to reuse (entrained) disfavored terms in Session 2 than participants who named the objects spontaneously during the norming task (4%[7%]; V=0, p<.0001). As in Experiment 1, maintenance was not perfect: Participants were significantly less likely to reuse a disfavored term in Session 2 than to initially use them in Session 1 (V=0, p<.0001).

For the items on which participants used the disfavored term in Session 1, they maintained disfavored terms on 68%[28%] of trials. Figure 6.5 illustrates how this rate varied depending on the speech communities of the two partners. As in Experiment 1, there was no significant effect of Second Partner's Community (β=.004, SE=.14, z=0.03, p>.05, see Table 6.6): Participants maintained disfavored terms at

similar rates across two partners from different communities (M=68%, SD=27%) as across two partners from the same community as each other (M=69%, SD=29%). Unlike Experiment 1, we did not find a significant effect of First Partner's Community (β=-.01, SE=.14, z=-.1, p>.05, see Table 6.6): Participants maintained disfavored terms at similar rates when the first partner was an in-community partner (M=69%, SD=28%) and when they were an out-community partner (M=68%, SD=29%).

Critically, however, we replicated the significant interaction between First Partner's Community and Second Partner's Community (β=-.3, SE=.13, z=-2.3, p=.02, see Table 6.6). That is, participants' tendency to maintain an entrained term was modulated by the second partner's speech community relative to the first partner's speech community.

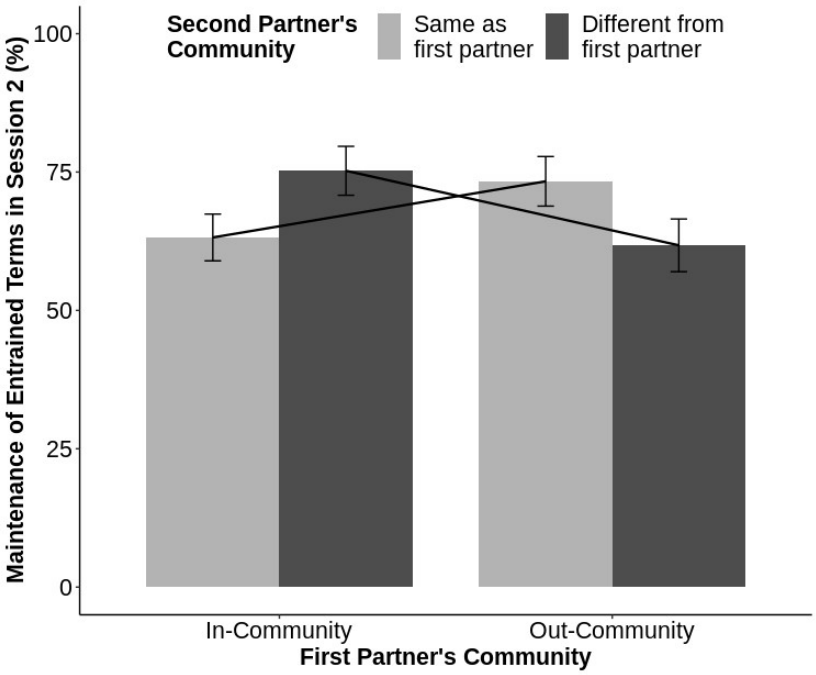


**Figure 6.5.** Mean and standard error of percentage of (re)use of disfavored terms in Session 2 (y-axis) across First Partner's Community (x-axis) and Second Partner's Community (color-coded). Solid lines represent a significant interaction between the First Partner's Community and Second Partner's Community on maintenance of entrained terms.

*Table 6.6. Experiment 2 – Session 2 Trials: Maintenance ~ FirstPartner * SecondPartner + (1|participants) + (1+ FirstPartner * SecondPartner |items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | 0.88822 | 0.23903 | 3.716 | **0.000202** |
| First Partner's Community (First) | -0.04111 | 0.14747 | -0.279 | 0.780437 |
| Second Partner's Community (Second) | -0.01154 | 0.15600 | -0.074 | 0.941053 |
| First:Second | -0.50364 | 0.16295 | -3.019 | **0.001997** |

To understand this interaction, we subset our data based on whether participants had entrained to an in-community partner or out-community partner, and then tested the effect of Second Partner's Community. We found a significant, positive effect for Second Partner's Community on in-community partner trials. In particular, participants who first entrained to an in-community partner subsequently maintained those terms significantly more often with an out-community partner (75%[27%]) than with another in-community partner (63%[26%]; β=0.45, SE=.18, z=2.5, p=.012, see Table 6.7 and Figure 6.5). Moreover, and replicating Experiment 1, there was a significant, negative effect of Second Partner's Community on out-community partner trials. In particular, participants who first entrained to an out-community partner subsequently maintained those entrained terms less often with an in-community partner (62%[29%]) than with another out-community partner (73%[27%]; β=-.49, SE=.21, z=-2.3, p=.021, see Table 6.8 and Figure 6.4).

*Table 6.7. Experiment 2 - In-Community Partner Trials: Maintenance ~ SecondPartner + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | 0.8900 | 0.2352 | 3.785 | **0.00154** |
| Second Partner's Community | 0.4497 | 0.1796 | 2.504 | **0.012297** |

*Table 6.8. Experiment 2 - Out-Community Partner Trials: Maintenance ~ Second-Partner + (1|participants) + (1 + SecondPartner |items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | 0.8830 | 0.3118 | 2.832 | **0.00463** |
| Second Partner's Community | -0.5165 | 0.2585 | -1.998 | **0.04572** |

Overall, these results confirm Experiment 1's findings that speakers maintain entrained terms across partners, and moreover that this tendency is mediated by the speech community membership of those partners. More specifically, as in Experiment 1, they suggest that speakers extrapolate more often from individual encounters with an out-community partner to other out-community partners from the same community than to in-community partners (i.e., from the speaker's own community). They also suggest that speakers do not extrapolate more often from an in-community partner to another in-community partner than to an out-community partner. However, unlike Experiment 1, Experiment 2's results suggest that speakers extrapolate lexical knowledge from individual linguistic encounters with an in-community partner to an out-community partner more often than to another in-community member.

Experiment 2 therefore replicates Experiment 1 in finding no evidence that speakers' beliefs about a partner's speech community affect their likelihood of entraining on the same term, but suggesting that they affect speakers' tendency to maintain an entrained term in a subsequent interaction with a new partner, in turn suggesting that speakers encode community information about their partner during lexical processing.

Critically, Experiment 2's replication of Experiment 1's key findings indicates that both the absence of effects of a partner's speech community on lexical entrainment, and the contrasting presence of effects of partners' speech community on maintenance of entrained terms, generalise across speakers from different speech communities. In particular, neither Castilians (Experiment 1) nor Mexicans (Experiment 2) showed more entrainment to out-community partners than in-community partners. Moreover, both Castilians (Experiments 1) and Mexicans (Experiment 2)

maintained out-community partners' terms more often with another partner from the same out-community than with an in-community partner, but did not maintain in-community partners' terms more often with another in-community partner than with an out-community partner (see Figures 3 and 5).

Experiment 2 ruled out that Experiment 1's findings followed from speakers' expectations about out-community partners' knowledge of in-community preferences. However, it is possible that our experimental design, and hence results, were confounded by linguistic status. In both Experiments 1 and 2, the in-community partner versus out-community partner manipulation correlated with a hierarchy for language variety status: Previous studies have suggested that Castilians (Experiment 1 in-community partners) perceive Latin American varieties of Spanish (spoken by Experiment 1 out-community partners) more negatively than Castilian Spanish, and Mexicans (Experiment 2 in-community partners) perceive Argentinian Spanish (spoken by Experiment 2 out-community partners) more negatively than the Mexican variety Spanish (Moretti, 2014; Chiquito & Quesada Pacheco, 2014). Given previous evidence that linguistic attitudes and socio-historical factors such as linguistic status can affect speakers' tendency to accommodate their partners' language use during dialogue (e.g., Gregory & Webster, 1996; Thakerar, Giles, & Cheshire, 1982; Nettle, 1999; Gallois & Callan, 1991; Palomares, Giles, Soliz, & Gallois, 2016), it is possible that our findings reflect not only the effect of beliefs about speech community membership, but also of linguistic status. Experiment 3 therefore set out to discriminate any effect of linguistic status on speakers' tendency to maintain out-community partners' terms more often with other out-community partners than with community members.

## 6.4. Experiment 3: Are community effects confounded with linguistic status?

Experiment 3 tested whether maintenance of entrained terms was influenced by the linguistic status of the out-community language variety, drawing on recent work indicating that Mexican speakers perceive their linguistic variety less positively

than Castilian Spanish but more positively than Argentinian Spanish (Morett, 2014). Thus, we compared Mexican participants' tendency to entrain to a (higher-status) Castilian out-community partner versus a (lower-status) Argentinian out-community partner, and how they maintained terms when subsequently interacting with a Mexican community member. If the results of Experiments 1 and 2 reflect participants' beliefs about linguistic status rather than speech community, then Mexican participants should be more likely to maintain disfavored terms introduced by high-status Castilian partners than by lower-status Argentinian partners.

### 6.4.1. Method

Unless otherwise detailed, the procedure was as in Experiments 1 and 2.

**Participants.** We recruited 40 native speakers of Mexican Spanish, aged 18-50, born and raised in Mexico.

**Items.** Across both sessions, we used the same items as in Experiment 2.

**Design.** We used a 2 (First Partner's Status: *Higher-Status* versus *Lower-Status*) between-participants and within-items factorial design. The dependent variables were (a) Lexical Entrainment (i.e., participants' use of disfavored terms in Session 1) and (b) Maintenance of Entrained Terms (i.e., participants' reuse of their own lexical choice from Session 1 in Session 2).

**Materials and Procedure.** We used the same materials as in Experiment 2, and the structure of the matching and naming task was the same as in Experiments 1 and 2 (Figure 6.1B). In the first session, we manipulated participants' beliefs about whether their partner was a member of a lower-status speech community (i.e., Agentina) or a higher-status speech community (i.e., Spain). All participants were told they would interact with an in-community partner from their own middle-status speech community (i.e., Mexico) in the second session.

### 6.4.2. Results

**Data processing and exclusions.** Coding and exclusions were carried out as in Experiments 1 and 2. Trials where participants named or selected the distractor

instead of the item picture were coded as NA. 2 participants reported believing that they had not played with a real person.

**Lexical Entrainment.** As in Experiments 1 and 2, in Session 1 participants used the disfavored term on approximately half of the trials (55%[32%] by-participants and 55%[11%] by-items). The frequency of use of the disfavored terms was significantly higher than in the norming task (4%[6%]; V=0, p<.0001), clearly suggesting the presence of a Lexical Entrainment Effect (see Figure 6.6).

However, we found no significant effect of Partner's Community on Lexical Entrainment (β=.09, SE=.24, z=.38, p>.05, see Table 6.9). Participants were as likely to entrain to a partner's use of a disfavored term when they believed their partner was from a higher-status speech community (i.e., Spain; 56%[34%]) as when they believed their partner was from a lower-status speech community (i.e., Argentina; 52%[31%]). A comparative analysis indicated that a null model, i.e., with only a fixed intercept, fit the data slightly better than a model regressing lexical entrainment against First Partner's Community, by a Bayes Factor of e(1609.85 – 1602.64)/2 = 36.74, and a posterior probability in favour of the null model BF / (BF + 1) = .97, which represents strong evidence against the hypothesis that speakers entrain to Higher-Status partners to a different extent than to Lower-Status partners.

*Table 6.9. Experiment 3 - Session 1 Trials: LexicalEntrainment ~ FirstPartnerStatus + (1|participants) + (1|items)*

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | 0.39079 | 0.28639 | 1.365 | 0.172 |
| First Partner's Status | 0.09018 | 0.23951 | 0.377 | 0.707 |

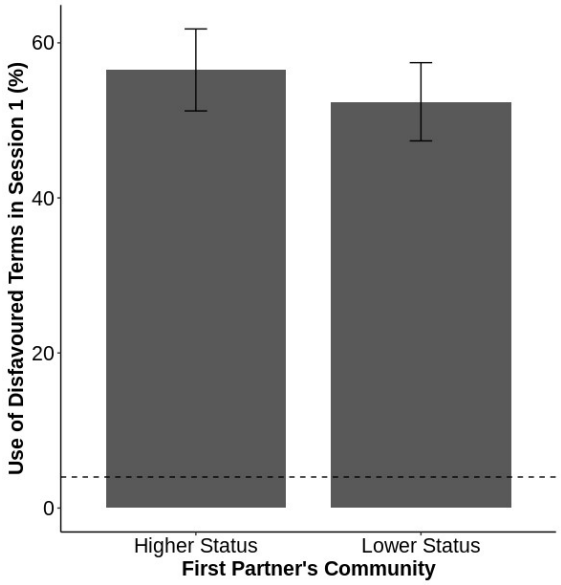


**Figure 6.6.** Mean and standard error of the percentage of use of disfavored terms in Session 1 (y-axis) across First Partner's Community (x-axis). The horizontal dashed line represents the mean of percentage of use of disfavored terms on the pretest.

**Maintenance of Entrained Terms.** As in Experiments 1 and 2, in Session 2 participants used disfavored terms on approximately one third of trials (28%[10%] of naming trials). Importantly, participants tended to maintain entrained terms: Participants reused (entrained) disfavored terms to name an object significantly more often than participants who named the objects spontaneously in the norming task (4%[6%]; V=0, p<.0001). In particular, after using a disfavored term in Session 1, participants maintained that disfavored term 68%[26%] of the time.

However, there was no significant effect of First Partner's Community (β=.24, SE=.23, z=1.02, p>.05, see Table 6.10 and Figure 6.7). Participants maintained a disfavored term they had previously used with a partner from a lower-status community in Session 1 (62%[29%]) as often as a disfavored term they had used with a partner from a higher-status community in Session 1 (56%[33%]). In other words, we found no evidence for an effect of speech community's linguistic status on participants' tendency to maintain an entrained disfavored term from a partner from another community compared to a partner from their own community.

Table 6.10. Experiment 3 – Session 2 Trials: Maintenance ~ FirstPartnerStatus + (1|participants) + (1|items)

| **Predictors** | **Estimate (β)** | **Std. Error (SE)** | **z-value** | **P-value** |
|---|---|---|---|---|
| (Intercept) | 0.1493 | 0.3893 | 0.384 | 0.701 |
| First Partner's Status | 0.2424 | 0.2374 | 1.021 | 0.307 |

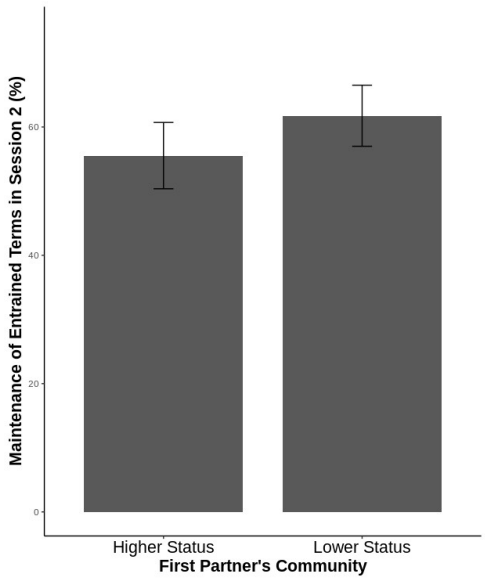


**Figure 6.7.** Mean and standard error of percentage of maintenance of disfavored terms in Session 2 (y-axis) across First Partner's Community (x-axis).

Overall, Experiment 3 found no evidence that speakers' beliefs about a partner's speech community status affected their likelihood of entraining on the same term for an object and, most importantly, did not suggest that such beliefs affect speakers' tendency to maintain an entrained term in a subsequent interaction with a partner from their own speech community. Critically, the null effect of linguistic status on maintenance suggests that Experiment 1 and 2's results do not reflect effects of linguistic status, but rather effects of beliefs about a partner's speech community membership.

## 6.5. Discussion

We know that speakers' referential choices are influenced by the personal history they share with their interlocutor, but we know very little about how the referential expressions that speakers use with a particular partner might be influenced by their beliefs about their partner's speech community, and by their previous history of interaction with other members of that partner's community. We investigated the interplay between interpersonal- and community-level influences on speakers' referential expressions, by examining how the speech community membership of a partner affects speakers' tendency to reuse a partner's referential expressions during interaction, and also their tendency to extrapolate those referential expressions to a subsequent partner from the same community as the first one.

In our three experiments, participants completed two sessions of a matching and naming task, swapping partners between sessions. In Session 1, we measured how a partner's use of a disfavored term for a picture (e.g., *papa* [potato]) influenced participants' subsequent use of that disfavored term. In Session 2, we measured whether participants would maintain the entrained disfavored term with a new partner (who never named the picture). Our critical questions concerned how lexical entrainment and referential maintenance were affected by the purported speech communities of the two partners.

In Experiment 1, our participants were Castilian and we induced them to believe that their partners were either also Castilian, i.e., a partner from their own speech community, or Latin American, i.e., a partner from a different community. In Session 1, participants showed significant lexical entrainment, but this was not critically affected by speech community: Use of disfavored terms was elicited to a similar degree by a partner from their own community as a partner from another community. In Session 2, participants tended to maintain these entrained terms, although at a lower rate than in Session 1. Interestingly, however, maintenance in Session 2 was signficantly affected by the new partner's speech community (relative to the first partner's speech community). In particular, when participants had entrained to an out-community partner (i.e., a partner from another community), they

maintained those terms more often with a new partner from the same out-community than with a new partner from their own community. But when participants had entrained to an in-community partner (i.e., a partner from their own community), they did not maintain entrained terms more often with other in-community partners than with out-community partners.

Experiments 2 and 3 replicated and extended these findings. In Experiment 2, we found a similar pattern of results with Mexican participants who believed that their partners were either Mexican in-community partners or Argentinian out-community partners. Replicating Experiment 1, lexical entrainment was not affected by partner's community, but maintenance of entrained terms in the second session was critically affected by the new partner's relative speech community. Importantly, Experiment 3 showed that the differential effects of speaker community on maintenance could not be explained in terms of linguistic prestige (e.g., participants being less willing to maintain terms introduced by a speaker from a lower-status community when interacting with a partner from their own community). In this experiment, Mexican participants entrained to partners from other speech communities that were either lower-status (Argentina) or higher-status (Spain), and then interacted with a Mexican in-community partner in Session 2. Linguistic status affected neither entrainment nor maintenance, suggesting that the effects in Experiments 1 and 2 were driven by differences in beliefs about the speech community of a partner (same as first partner versus different from first partner), rather than differences in the linguistic status of a partner's speech community (high- versus low-status).

Taken together, our experiments replicate and extend to Spanish-speaking populations previous results showing that lexical entrainment occurs even when not interacting face-to-face with a partner (e.g., Branigan et al., 2011; Tobar-Henríquez, Rabagliati & Branigan, 2019). Furthermore, they provide novel evidence that speakers' tendency to maintain previously used referential expressions when interacting with new partners similarly occurs in such computer-mediated interactions.

But most importantly, our results provide new insights into how we learn community-appropriate language from individual encounters with particular partners. They suggest an interesting dissociation in how the speech community of a partner affects lexical entrainment versus maintenance of entrained terms. Across our three studies, entrainment was never affected by speech community membership. But maintenance of entrained terms *was* affected by the relative identity of the new partner's speech community (Experiments 1 and 2), although not by a partner's community status (Experiment 3). These positive maintenance results are particularly important: They indicate that, even though a partner's community did not affect the degree to which participants entrained to that partner, participants' beliefs about their partner's community were nevertheless incorporated during the linguistic interaction in which they entrained, and were stored alongside each entrained term to inform, or constrain, its future contexts of use.

This pattern suggests that, just as participants build interpersonal-level common ground during particular interactions, they are also in parallel updating their knowledge of community-level common ground with respect to their partner's community – in other words, their knowledge of the linguistic preferences shared by members of their partner's community. Moreover, the dissociation between initial entrainment and subsequent maintenance indicates that the creation of interpersonal-level common ground and knowledge of community-level common ground can also be dissociated, suggesting that community-level norms are more than just the sum of our experiences with individuals.

Our lexical entrainment effects speak to mediated versus unmediated theories of linguistic entrainment. In certain respects, the null findings regarding effects of community membership on lexical entrainment in Session 1 are consistent with unmediated accounts of entrainment (e.g., Pickering & Garrod, 2004), as rates of entrainment did not differ based on speech community membership (in-community partner versus out-community partner, Experiments 1 and 2) or partner's community status (higher- versus lower-status, Experiment 3). Thus, if the entrainment data from our studies were considered in isolation, they would be consistent with entrainment being a low-level phenomenon, perhaps reflecting basic memory mechanisms. But

the fact that beliefs about linguistic community affected subsequent maintenance suggests that the lexical use engendered by our entrainment manipulation did incorporate higher-level information about beliefs, which is not consistent with unmediated accounts of entrainment. How can this tension be resolved?

One possibility is that our manipulation of beliefs in these experiments, concerning membership of different communities, is not the type of belief manipulation that critically affects entrainment at the level of an individual interaction. Under this possibility, entrainment would be mediated by beliefs but, in this particular task, beliefs about community membership would be overridden by beliefs about the most appropriate way to interact with this particular individual partner, given their apparent idiosyncratic preferences. For example, a speaker might use a low-frequency term like *papa* with a member of their own speech community, even though it is not the preferred community term, because their partner has made it plain by their previous usage that this is the term they individually prefer.

This account of entrainment has the advantage that it can potentially explain why our results differ from those of Branigan et al. (2011), who found that participants entrained more when they believed they were interacting with computer partners compared with human partners. In the introduction, we suggested that this pattern could have been an effect of linguistic community (with humans as in-community partners, and computers as out-community partners). But our results suggest that that is not the case. Instead, a more likely explanation, given our data, is that entrainment was stronger with computers because participants believed that they were interacting with entities whose comprehension skills were lower, and so maximised their use of entrained terms to enhance comprehension.

A second possible explanation for why beliefs were encoded during entrainment, but did not affect the likelihood of entrainment, is that the measurement properties of our task were not sensitive enough to capture the effects of beliefs on lexical entrainment. In particular, entrainment may be affected by both higher-level information such as beliefs *and* lower-level processes such as priming (Branigan et al., 2010; Branigan et al., 2011). However, our task may have maximised the influence of priming over the influence of beliefs. For example, since matching

(prime) trials and naming (target) trials were closely separated in time, we may have inadvertently caused participants to be at ceiling in using entrained terms.

Alternatively, it could be that our manipulation of beliefs was too weak. For example, community membership may affect entrainment only in more interactive settings where social components are more salient, such as in face-to-face interaction. We note that studying entrainment in more naturalistic settings such as face-to-face interactions presents challenges in interpreting effects, as enhanced entrainment in such contexts could be caused not only by beliefs about speech community, but also factors such as increased interactivity, differences in language use exhibited by members of different speech communities, individual speakers' appearance, etc. The computer-mediated internet-based method that we used allowed us to directly test the effects of beliefs about a partner on entrainment and maintenance, independently of a partner's behaviour.

We note that studying entrainment in more naturalistic settings presents challenges in interpreting effects, as enhanced entrainment in such contexts could be caused not only by beliefs about speech community, but also factors such as increased interactivity, differences in language use exhibited by members of different speech communities, individual speakers' appearance, etc. Thus, there is still significant work to be done in understanding whether and how beliefs about a partner's speech community affect lexical entrainment. But what is clear from our data is that, first, beliefs about community do not inevitably affect entrainment; and second, even when beliefs about community have no observable effects on entrainment, such beliefs are nonetheless tracked during interactions, and associated with lexical representations, in ways that affect subsequent language use.

Why did beliefs about partner's community affect maintenance of terms, even when they did not affect entrainment to terms during the first (entrainment) session? We suggest that these results follow because speakers' audience design is sensitive to their beliefs about, and experience with, conventions of different speech communities, and because speakers are more certain about the conventions of their own community than about the conventions of other communities in which they are not members. More specifically, we suggest that participants showed a tendency to

extrapolate out-community partners' disfavored terms to other partners from the same out-community because their knowledge about their out-community partner's community preferences was insufficiently detailed to reliably distinguish individual preferences from community preferences. Therefore, an out-community partner's use of a disfavored term acted as positive evidence that members from that community would prefer the disfavored term, updating participants' knowledge of that community's preferences and increasing their likelihood to maintain that term with a subsequent partner from the same community.

In contrast, participants did not extrapolate in-community partners' terms more often to other in-community partners than to out-community partners, because their knowledge of their own community's preferences included the information needed to distinguish community preferences from individual preferences. That is, they had a solid understanding of their own community's preferences, and thus a new linguistic interaction where an in-community partner used a disfavored term did not critically update participants' knowledge about their own community's preferences, leaving unchanged their likelihood of maintaining its precedent. Importantly, this interpretation is consistent with previous work showing that the more linguistic experience speakers have, the less likely they are to update their linguistic knowledge based on new linguistic encounters (Lev-Ari, 2018).

Our data are also relevant to previous proposals that lexical entrainment (and thus perhaps maintenance) are influenced by linguistic status (e.g., Thakerar et al., 1982; Nettle, 1999). These proposals offered an alternative interpretation for Experiments 1 and 2's results, because the 'out' speech communities that we used might have been perceived as less prestigious than participants' own speech communities (Chiquito & Quesada Pacheco, 2014). Interestingly, however, entrainment in our studies was not affected by linguistic status, suggesting that lexical entrainment in computerised interaction is not sensitive to partner's linguistic status. One possibility is that previously reported effects of linguistic status on entrainment were specific to the particular speech communities studied, and do not generalise to Spanish-speaking communities. It is also possible that our paradigm was not sensitive enough to capture the effects of linguistic status on lexical

entrainment. Alternatively, it is possible that previous results do not replicate in our studies because previous studies were confounded with other factors, in the same ways as discussed above (e.g., differences in language use exhibited by members of different communities, degree of interactivity of individual encounters, differences in speakers' degree of affiliative behaviours [e.g., van Baaren, Holland, Steenaert, & van Knippenberg, 2003]). But importantly, we have shown that, at least in our paradigm, lexical adaptation to a partner's speech community does not reflect the effects of linguistic status.

Moreover, our data also suggested that linguistic status did not affect maintenance. In principle, the maintenance effects found in Experiments 1 and 2 might have occurred because participants were more willing to extrapolate (lower-status) entrained terms used with a partner from another community more often to other partners from the same out-community than to members from their own community because of the potential social cost associated with using a lower-status community preference (Palomares et al., 2016; Thakerar et al., 1982), rather than due to how confident they were about their partner's community preferences. However, Experiment 3 suggested that variations in status do not affect maintenance to a significant degree, when status is decoupled from shared community membership. Specifically, participants extrapolated to other members of their own speech community entrained terms from a lower-status partner as often as those from a higher-status partner, supporting the conclusion that results from Experiments 1 and 2 do not reflect linguistic status effects.

As predicted, in both Experiments 1 and 2 participants maintained terms they had entrained on with in-community partners to the same extent with subsequent in-community partners as out-community partners. However, there was a minor unexplained difference between the experiments: While participants in Experiment 1 maintained in-community partners' terms equally often with in-community partners and out-community partners, participants in Experiment 2 maintained in-community partners' terms less often with in-community partners than out-community partners. In other words, Experiment 1's (Castilian) participants did not distinguish between in-community partners and out-community partners when extrapolating disfavored

terms, whereas Experiment 2's (Mexican) participants extrapolated disfavored terms less often to in-community partners than out-community partners.

Why this happened is uncertain, but an interesting possibility is that this difference reflects historical differences between Spain and Latin America, and Mexico and Argentina. In particular, due to the Spanish colonisation of the Americas, Castilians may assume that Latin Americans are familiar with Castilian lexical preferences (Castilian favored terms), thus leading Castilians to maintain disfavored terms to similar extents with Latin Americans and Castilians (see Chávez Fajardo, 2014, for discussion of linguistic homogenisation in the standardisation processes of Latin American varieties of Spanish). Mexicans, however, do not necessarily have good reasons to assume that Argentinians are familiar with Mexican lexical preferences: There is no clear historical supremacy of the Mexican variety of Spanish over the Argentinian variety, and Mexicans are likely to be aware that Latin American varieties of Spanish differ in lexical preferences (though not necessarily knowing details of these preferences). Mexicans may thus have attempted to neutralise their own Mexican preferences in favour of using a more neutral, standard Spanish with Argentinians, by maintaining disfavored Mexican terms more often with Argentinians versus Mexicans. This interpretation, though speculative, is consistent with previous social-historical accounts of lexical accommodation during inter-cultural interactions, highlighting the importance of cultural identity and historical factors in language use (e.g., Communication Accommodation Theory; Giles et al., 2003).

In sum, our three experiments showed how community membership does, and does not, affect speakers' referential expressions during dialogue. We found that entrainment to a partner's lexical choices was not affected by their purported speech community. But speech community did affect whether participants maintained these entrained terms with a new partner, with participants showing less maintenance of disfavoured terms when they moved from an out-community partner to an in-community partner. Importantly, this latter effect appeared to be specific to beliefs about a partner's membership of another speech community, rather than any associated status judgments about that community. These findings suggest that when

we interact with a partner, we not only encode what they say, but also information about their speech community. This additional contextual representation might not affect usage within the interaction, specifically likelihood of reusing their term, but does affect usage in subsequent interactions. Thus, interaction in individual level contexts leads not only to the creation of individual common ground, but also knowledge of community-level common ground, so that we extrapolate language learned in interpersonal-level contexts to broader community-level contexts.

Chapter 7

# 7. Conclusions

## 7.1. Introduction

This thesis aimed at better understanding the variability of language use by examining individual-, interpersonal-, and community-level influences on speakers' tendency to vary their referential choices. In particular, we focused on the tendency for a speaker to reuse their interlocutor's referential expressions (lexical entrainment; e.g., using *brolly* after your partner used *brolly*), and to generalise an entrained expression across two interlocutors (referential maintenance; e.g., using *brolly* after you entrained to your partner's use of *brolly*).

As discussed in Chapter 2, it is uncontroversial that people tend to reuse their partners' words during interaction, but the mechanisms that underpin this are still a matter of debate. Unmediated theories of lexical entrainment suggest that speakers reuse a partner's words due to recent linguistic processing that is not mediated by beliefs (Pickering & Garrod, 2004). By contrast, mediated theories of entrainment explain the phenomenon as a result of lexical processing being mediated by speakers' beliefs: audience design accounts suggest entrainment occurs because speakers adopt their partner's perspective to facilitate mutual comprehension (Clark, 1996), and social affiliation accounts suggest that speakers entrain as a way to increase social affiliation with their interlocutor (Van Baaren, Holland, Steenaert, van Knippenberg, 2003). Since lexical entrainment can lead speakers to deviate from their community-level norms (e.g., using the disfavoured label *brolly* instead of the favoured label *umbrella* in British English), and given that lexical entrainment mechanisms - lexical retrieval, audience design, and social affiliation - can vary widely across both

individuals and communicative situations, this phenomenon is an excellent test-case to understand the interplay of individual-, interpersonal-, and community-level influences in how and why speakers vary their lexical choices. In four studies, we conducted individual differences studies and group-level comparisons to understand individual-, interpersonal-, and community-level influences on lexical entrainment and maintenance of entrained terms.

At the time we started this investigation, it was already uncontroversial that lexical entrainment had a high experimental reliability, and thus that we could use group-level comparisons to examine the role of situational factors in lexical entrainment. However, it was not clear whether lexical entrainment tasks were also reliable for individual differences studies. We thus started by understanding the extent to which a lexical entrainment task could reliably capture individual variation. After having demonstrated that this was the case, we used the task to investigate individual differences in speakers' tendency to entrain. Then, we moved on to using experimental manipulations to understand the effects of situational factors on lexical entrainment with a partner, and on maintenance of entrained referential precedents with subsequent partners.

## 7.2. Overview of studies and findings

The first study (Chapter 3) presented a web-based adaptation of a well-established lexical entrainment task (i.e., Branigan et al., 2011), and investigated a crucial psychometric property for the study of individual differences, the task's test-retest reliability. In two studies, native speakers of British English completed two sessions of a matching-and-naming task, where they took turns with a partner to first match and then name an object that had both a favoured name (e.g., *umbrella*) and a disfavoured but highly acceptable name (e.g., *brolly*) in British English. Participants always experienced their partner naming critical targets with disfavoured names (e.g., *brolly*) before themselves naming the targets, and lexical entrainment was then measured as participants' use of disfavoured names. In Experiment 1, the two

sessions were separated by 2 minutes; in Experiment 2, they were separated by a week. In each experiment, we examined the test-retest reliability of lexical entrainment by calculating the consistency between participants' tendency to entrain in the first session and their tendency to entrain in the second session.

Overall, participants tended to entrain to their partner, supporting the previously attested experimental reliability of the task. More importantly for individual differences research, the degree of lexical entrainment varied widely across individuals, and their tendency to entrain was consistent both across two sessions separated by minutes and across two sessions separated by a week. Taken together, these results suggest that our lexical entrainment task reliably captures individual differences and can then be used to understand individual-level influences in the propensity to entrain.

The second study (Chapter 4) examined the extent to which lexical entrainment was predicted by individual differences in schizotypy and age, which have both been linked to divergent functioning of potential lexical entrainment mechanisms (lexical retrieval, audience design, and social affiliation). Online native speakers of British English (aged 26-65) engaged in one session of a matching-and-naming task, where we used materials normed for this particular population.

We again found that participants tended to reuse their partner's terms. Although schizotypy did not predict individuals' tendency to entrain, age did predict entrainment: The older participants were, the more often they reused their partner's terms. These results suggest that individual differences in lexical entrainment are not random, and that at least some lexical entrainment mechanisms may vary across the lifespan. However, ageing correlates both with increased social affiliation and difficulties in lexical retrieval, and it is thus unclear which of these mechanisms underlie the positive relationship between entrainment and age.

The third study (Chapter 5) investigated the effects of social affiliation in lexical entrainment, by combining individual differences and group-comparison approaches. Previous research has shown that experiencing ostracism increases affiliative behaviours as a compensatory measure to regain social acceptance; thus, if lexical entrainment involves an affiliative component, then experiencing ostracism

should lead to increased lexical entrainment. Importantly, this is a causal relationship that has been suggested by previous research, but that had not been empirically tested. Moreover, it is unclear if increased affiliative dispositions are targeted to repairing individual relationships or to enhancing social affiliation more generally, and it is also uncertain whether ostracism-induced increased affiliative dispositions are moderated by individual differences in basal social affiliation dispositions. In two identical experiments, native speakers of British English engaged in two web-based interactive tasks: They first played a ball-tossing game with two partners, who either ostracised them from the game or did not ostracise them; they then completed our lexical entrainment task, with either a partner from the previous ball-tossing game or a new partner. Finally, we measured participants' neuroticism scores.

As in the previous two studies, participants tended to reuse their partner's terms. Critically, in Experiment 1, ostracised participants were more likely to entrain than non-ostracised participants, but they entrained as often with a partner who had ostracised them as with an 'innocent' partner; moreover, post-ostracism entrainment was not predicted by neuroticism. Experiment 2 replicated the pattern that ostracised participants entrained more often than non-ostracised participants, but this difference did not reach statistical significance; however, Experiment 2 suggested that post-ostracism lexical entrainment was moderated by neuroticism. A combined analysis of both experiments suggested that ostracism increased the likelihood of entrainment, and that this effect was not moderated by their previous interaction with their partner, suggesting that our mixed results concerning the effect of ostracism on entrainment might have emerged from statistical power differences across the experiments. Importantly, our combined analyses supported Experiment 2's result that post-ostracism lexical entrainment was indeed moderated by neuroticism, suggesting that our mixed results concerning neuroticism may have also resulted from differences in statistical power. Taken together, these results suggest that entrainment implies an affiliation component, that such a component seems to be targeted to enhancing social affiliation in general (rather than repairing individual relationships), and that compensatory affiliative linguistic behaviours seem to be moderated by individual-level influences.

The fourth study (Chapter 6) moved beyond understanding only interpersonal-level influences on speakers' lexical choices, to examine whether and how speakers' choices are affected by community-level influences, such as whether your partner is from your own community (*in-community partner*) or from another community (*out-community partner*). In three experiments, participants completed two sessions of a matching-and-naming task, swapping partners between sessions. In Session 1, we measured lexical entrainment to their partner. In Session 2, we measured whether participants would maintain the entrained disfavored term with a new partner (who never named critical targets).

In Experiment 1, our participants were Castilian and we induced them to believe that their partners were either also Castilian, i.e., a partner from their own speech community, or Latin American, i.e., a partner from a different community. In Session 1, participants tended to entrain to their partner, but they entrained as often with an in-community partner as with an out-community partner. In Session 2, these entrained terms tended to be maintained, and maintenance was significantly affected by the new partner's speech community (relative to the first partner's speech community). In particular, when participants had entrained to an out-community partner (i.e., a partner from another community), they maintained those terms more often with another out-community partner than with a partner from their own community. But when participants had entrained to an in-community partner (i.e., a partner from their own community), they did not maintain entrained terms more often with other in-community partners than with out-community partners. Experiments 2 and 3 replicated and extended these findings. In Experiment 2, we found a similar pattern of results, but using Mexican participants who believed that their partners were either Mexican in-community partners or Argentinian out-community partners. Importantly, Experiment 3 showed that the differential effects of speaker community on maintenance could not be explained in terms of linguistic prestige, suggesting that the effects in Experiments 1 and 2 were driven by differences in beliefs about the speech community of a partner (same as first partner versus different from first partner), rather than differences in the linguistic status of a partner's speech community (high- versus low-status). Taken together, these results suggest that

lexical information interacts with social information during lexical entrainment, such that beliefs about a partner's community (and their likely lexical preferences) are tracked during interaction, and stored along with lexical representations, to inform future contexts of use. This in turn suggests that community-level influences affect linguistic processing during lexical entrainment.

## 7.3. Implications for accounts of lexical entrainment and future research

Taken together, our findings suggest that lexical entrainment can be affected by the interplay between linguistic and non-linguistic processing, and that how this interplay is realised depends on individual-, interpersonal-, and community-level influences. This in turn suggests that theories of lexical entrainment need to keep moving away from the unmediated linguistic processing versus mediated linguistic processing dichotomy, to instead focus on understanding which cognitive mechanisms shape how linguistic and non-linguistic information interact during entrainment, and how such mechanisms enable variation at the individual, interpersonal, and community levels.

In particular, the findings that lexical entrainment is (i) stable within individuals, (ii) increases with age, and (iii) can be moderated by individual differences in personality, demonstrate that speakers' likelihood to exhibit this behaviour is sensitive to individual-level influences. Moreover, evidence that lexical entrainment increases as a result of the emotional distress inflicted by ostracism demonstrates both that (i) lexical and socio-affective information interact during entrainment, and (ii) lexical entrainment is sensitive to interpersonal-level influences. Critically, the finding that the effects of ostracism on entrainment is moderated by individual differences in personality suggests that the effects of interpersonal-level influences interact with individual-level influences. Furthermore, the fact that speakers generalise entrained expressions across interlocutors based on their speech communities suggests both that (i) lexical and social information interact during

entrainment, and that (ii) linguistic processing during entrainment is affected by community-level influences. In particular, the fact that speakers generalised entrained terms depending on their previous experience with their interlocutors' community-level preferences shows that interpersonal-level and community-level influences can interact with each other during lexical entrainment, to affect future language processing.

The fact that lexical entrainment reflects stable individual variation, along with the positive relationship between entrainment and age, offers an opportunity to understand better which cognitive mechanisms are involved in entrainment, and how their interaction shapes the variability of referential expressions from person to person. Given that older adults experience more difficulties than younger adults to retrieve lexical labels, it is possible that their increased likelihood to entrain to recently processed disfavoured labels resulted from difficulties in accessing the lexical representation of the corresponding favoured labels. Moreover, based on previous research suggesting that an increased tendency to entrain in children might be explained by low inhibitory control and other research suggesting decreased inhibition in healthy ageing, it is plausible that older participants entrained more often due to difficulties in inhibiting recently processed disfavoured labels. Naturally, it is also possible that difficulties to inhibit recently processed disfavoured labels also increased difficulties in retrieving favoured labels, thus increasing the likelihood of entrainment as a result from an interaction between inhibitory control and lexical retrieval skills. Alternatively, and based on previous evidence that older adults are more eager than younger adults to pursue social affiliation, it is possible that older participants' increased entrainment resulted from an increased interest in social affiliation.

Future research can thus further our understanding of lexical entrainment by examining its relationship with lexical retrieval, inhibition, and social affiliation dispositions. In particular, if the relationship between age and entrainment is explained by lexical retrieval deficits, such that older speakers entrained to disfavoured labels more often than younger speakers because they struggled to retrieve favoured labels from memory, then we should find that lexical entrainment

correlates with naming difficulties (e.g., Juncos-Rabadán, Facal, Rodríguez, & Pereiro, 2009). Similarly, if older participants entrained more often than younger adults because they struggled to inhibit recently processed disfavoured labels, then older individuals' tendency to entrain should be negatively predicted by inhibitory control (e.g., Brown-Schmidt, 2009). Moreover, if inhibitory control difficulties contributed to lexical retrieval difficulties, then we should find that participants with lower inhibition skills have lower lexical retrieval skills (as shown by more naming difficulties). In contrast, if the age effect we found is explained by social affiliation, such as older adults entrained more often than younger adults due to enhanced affiliation dispositions, then we should find that the degree of lexical entrainment is predicted by social affiliation disposition, such as the degree of cooperativeness in economic games (e.g., Matsumoto, Yamagishi, Li, & Kiyonari, 2016).

Critically for understanding how lexical and socio-affective information interact during lexical entrainment, we demonstrated novel evidence for the causal relationship between enhanced social affiliation dispositions and the propensity to reuse an interlocutor's referential expression. This opens new directions for investigating lexical entrainment mechanisms, and how the phenomenon is affected by both individual- and interpersonal-level influences. Since ostracism has been previously shown to increase social affiliative behaviours as a compensatory measure to regain social belonging (see Lakin et al., 2008), we interpret the fact that experiencing ostracism increased participants' likelihood of lexically entraining as evidence that the phenomenon implicates a social affiliation component. In other words, we suggest that ostracised participants entrained more often than non-ostracised participants as a way to enhance social affiliation with their partner, supporting both that lexical information interacts with socio-affective information during lexical entrainment, and that the likelihood of entrainment is affected by interpersonal-level influences. Strikingly, the effects of ostracism on speakers' increased likelihood to entrain was moderated by speakers' basal social affiliative dispositions; neuroticism, which is associated with a high need for social acceptance, positively predicted ostracised participants' tendency to entrain. This evidence not only strengthens the case for a social affiliation component in lexical entrainment,

but it also shows that the effects of interactional-level influences (e.g., being ostracised from a social group) on lexical entrainment can be moderated by individual-level influences (e.g., individuals' baseline social affiliative dispositions and needs).

Future research should now interrogate the cognitive mechanisms and socio-affective skills that make this interplay possible. For example, given previous research suggesting that ostracism increases non-conscious affiliative behaviours, the causal relationship between ostracism and lexical entrainment suggests that the tendency to reuse a partner's word may rely upon basic, non-conscious affiliation mechanisms (see Mead et al., 2011; Lakin et al., 2008). This idea is also consistent with the finding that ostracised participants entrained as often with a new partner, with whom they had not interacted before, as with a partner who had perpetrated ostracism towards them. This indicates that ostracism-induced increased lexical entrainment was targeted to increase social affiliation in general, supporting a basic, non-conscious component to their increased affiliative tendency. Thus, if lexical entrainment is underpinned by non-conscious social affiliation goals, it is possible that individual differences in lexical entrainment may correlate with basic social affiliation dispositions. For example, we should find that individuals' tendency to lexically entrain is predicted by non-conscious social affiliation behaviours, such as unconscious mimicry of emotional facial expressions (e.g., Mauersberger et al., 2015).

Relevant to the interplay between social and linguistic information during entrainment, the finding that participants generalised entrained terms based on their partner's community suggests that during lexical entrainment speakers can encode social information about their partner, store that information along with a lexical representation, and then use such social information to inform future contexts of language use. It is unclear why participants' tendency towards entrainment was not affected by their beliefs about their partner's community at an observable level, given that they evidently encoded such information. However, an interesting possibility is that they did not consider that their partner's community could restrict the likelihood of mutual comprehension if they did not entrain to their partner's use of disfavoured

terms. Under this logic, beliefs about community membership could have been overridden by beliefs about the most appropriate way to interact with a particular individual partner, given their apparent idiosyncratic lexical preferences. For example, a speaker might use a disfavoured label with a member of their own speech community because their partner has made it plain by their previous usage that that is the term they individually prefer. In principle, this result - that certain social beliefs are tracked during entrainment but they do not observably affect its likelihood - suggests that, during entrainment, speakers must be able to distinguish (albeit non-consciously) between non-linguistic information that is relevant and linguistic information that is not relevant to their communicative or social goals.

Critical to understanding how speakers identify relevant non-linguistic information during entrainment, our participants' tendency to transfer entrained terms across partners was influenced by their amount of previous experience with their partner's community-level lexical preferences, which suggests that long-term memory processes must have also been deployed during entrainment. In particular, participants generalised out-community entrained terms to out-community partners more often than to in-community partners, but this was not the case for in-community entrained terms. In other words, they were less prone to update their knowledge of their own community-level preferences, which they presumably knew very well, than their knowledge of another community's preferences, which they presumably did not know very well. This consideration suggests that, during lexical entrainment, participants accessed information from long-term memory and used that information to encode information about their partner's community. Therefore, future research could further our understanding of how speakers distinguish relevant and non-relevant non-linguistic information during entrainment, by examining individual differences in cognitive processed such as the abilities to maintain relevant information in working memory and make new memory connections.

In sum, our findings suggest that lexical entrainment is a complex phenomenon that is affected by the interaction between linguistic and non-linguistic information, is supported by multiple cognitive processes, and can be sensitive to individual-, interpersonal-, and even community-level influences.

## 7.4. Implications for the variability of language use

Our results provide insights into how the variability of language use is shaped by individual, interpersonal, and community factors, and by how these influences interact with each other. In particular, our individual differences studies in lexical entrainment demonstrate that the variability of language use can be influenced by individual-level factors. Moreover, the fact that speakers' lexical choices were affected by experiencing ostracism shows that language use can be sensitive to interpersonal-level influences, such as experiencing emotional distress during social interactions. Critically, the fact that the effects of ostracism on language use were moderated by individual differences in personality demonstrates that interpersonal-level influences on language use can be moderated by individual traits. Furthermore, the fact that speakers varied their referential expressions depending on their previous experience with their interlocutors' community-level preferences shows that language use can also be affected by the interplay between interpersonal-level and community-level influences.

More specifically, we have demonstrated that lexical entrainment increases with age and that it can be moderated by individual differences in the need for social acceptance, which suggests that speakers can, in principle, vary their language use depending on various individual-level factors, ranging from lexical retrieval and inhibition skills, to social affiliative dispositions and emotional needs. Critically, the finding that lexical entrainment reflects stable individual differences in language processing has promise to keep furthering our understanding of which individual-level factors determine the variability of language use, and how.

Moreover, we have demonstrated that the likelihood of lexical entrainment increases as a compensatory affiliative behaviour to regain social acceptance after experiencing ostracism, which suggests that speakers can vary their language use depending on their social dispositions and emotional needs. Given previous evidence

that ostracism increases non-conscious affiliative behaviours, such as behavioural mimicry, this result suggests that language use can be affected by non-conscious affiliative tendencies (e.g., Lakin et al., 2008), but which mechanisms underlie this interaction is still an open question. Importantly, however, we have shown that the effects of ostracism on speakers' language use were moderated by individual differences in neuroticism, which suggests that interpersonal-level and individual-level influences interact in shaping the variability of language use. Future studies could then keep furthering our knowledge of how interpersonal and individual influences interact, by combining group-comparisons and individual differences approaches.

In addition, we have shown that speakers' referential expressions are not only determined by individual differences or by interpersonal factors affecting a particular single interaction, but also by community-level influences and how they interact with interpersonal factors. In particular, we demonstrated that speakers generalise referential expressions across interlocutors depending on their beliefs about, and experience with, the lexical preferences of their interlocutors' speech communities. This evidence extends audience design accounts of language use; in particular, it suggests that the effects of community-level influences on reference seems to depend on the interplay between speakers' previous experience with their partner's community-level preferences (i.e., communal common ground) and speakers' recent positive evidence for such preferences (i.e., personal common ground). Critically, this evidence also supports that speakers extrapolate community-level knowledge from individual linguistic interactions, suggesting that community-level and interpersonal-level influences interact in how speakers vary their language use, and opening new research directions to understand which cognitive mechanisms enable the interplay between interpersonal and community-level influences.

In sum, our lexical entrainment and referential maintenance results suggest that the variability of referential expressions depends on individual, interpersonal, and community-level influences. We have shown that lexical entrainment is a good test-case for the study of individual differences in how speakers' vary their lexical choices, as it reveals stable individual variation; and we have learned that some such

individual differences in how speakers vary their lexical choices might relate to lexical retrieval skills and social affiliation dispositions and needs. Moreover, the variability of speakers' referential expressions is influenced by interpersonal influences, such as increased social affiliation goals and beliefs about a partner's likely community-level referential preferences. Critically, these interpersonal influences can interact with both individual-level influences (such as individual differences in social affiliative dispositions and needs) and community-level influences (such as knowledge of speech community's lexical preferences).

## 7.5. Conclusions

This thesis investigated individual, interpersonal, and community-level influences in the variability of language use through the lens of lexical entrainment, the tendency for a speaker to reuse the same word as their interlocutor. We found that lexical entrainment reflects stable individual differences in how speakers use language, suggesting both that the phenomenon has promise to keep furthering our knowledge of individual variation in language use, and that the sensitivity to a partner's referential choice varies across individuals. Moreover, we found that speakers vary their language use depending on interpersonal influences, such as their attunement to social affiliation and their beliefs about their partner's likely linguistic preferences. Furthermore, we have demonstrated that speakers' tendency to adapt their lexical choices during dialogue does not only depend on interpersonal factors affecting a particular single interaction, but also on community-level influences, such as speakers' previous experience with community-level linguistic preferences. Critically, we have also shown that interpersonal-level factors affecting language use can be moderated by both individual-level and community-level influences. Taken together, these results suggest that variation in language use emerges from the interplay between multiple sources, opening new directions for furthering our understanding of how language use works.

# 8. References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Andreasen, N. C. (1986). Scale for the assessment of thought, language, and communication (TLC). *Schizophrenia bulletin*, *12*(3), 473.

Arnon, I. (2019). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01205-5

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. Language in Society, 39(4), 437-456.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American psychologist*, *54*(7), 462.

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. Cognition, 109(1), 18-40.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Barresi, B. A., Nicholas, M., Tabor Connor, L., Obler, L. K., & Albert, M. L. (2000). Semantic degradation and lexical access in age-related naming failures. *Aging, Neuropsychology, and Cognition*, *7*(3), 169-178.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., & Green, P. (2015). "Package 'lme4'." Convergence 12, 1:2.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, *117*(3), 497.

Beadle, J. N., Sheehan, A. H., Dahlben, B., & Gutchess, A. H. (2015). Aging, empathy, and prosociality. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *70*(2), 213-222.

Bell, A. (1984). Language style as audience design. Language in society, 13(2), 145-204.

Berchtold, A. (2016). Test–retest: Agreement or reliability? *Methodological Innovations, 9*, pp. 1-7, https://doi.org/10.1177/205979911667287

Blanchard-Fields, F. (2007). Everyday problem solving and emotion: An adult developmental perspective. *Current Directions in Psychological Science*, *16*(1), 26-31.

Bleuler, E. (1950). *Dementia Praecox; Or, The Group of Schizophrenias: Or, The Group of Schizophrenias*. International Universities Press.

Bortfeld, H., & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non native speakers. Discourse Processes, 23(2), 119-147.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and speech*, *44*(2), 123-147.

Branigan, H. P., & Messenger, K. (2016). Consistent and cumulative effects of syntactic experience in children's sentence production: Evidence for error-based implicit learning. *Cognition*, *157*, 250-256.

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics*, 42(9), 2355-2368.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41-57.

Branigan, H. P., Tosi, A., & Gillespie-Smith, K. (2016). Spontaneous lexical alignment in children with an autistic spectrum disorder and their typically developing peers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1821-1831.

Branigan, H. P., Tosi, A., & Gillespie-Smith, K. (2016). Spontaneous lexical alignment in children with an autistic spectrum disorder and their typically developing peers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1821-1831.

Breen, W. E., & Kashdan, T. B. (2011). Anger suppression after imagined rejection among individuals with social anxiety. Journal of anxiety disorders, 25(7), 879-887.

Brennan, S. E. (1990). Conversation as direct manipulation: An iconoclastic view. The art of human-computer interface design, 393-404.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482-1493.

Brennan, S. E., & Hanna, J. E. (2009). Partner specific adaptation in dialog. Topics in Cognitive Science, 1(2), 274-291.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review, 16*(5), 893-900.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. Cognition, 107(3), 1122-1134.

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as contexts in conversation. *Psychology of learning and motivation* (Vol. 62, pp. 59-99). Academic Press.

Buckley, K. E., Winkel, R. E., & Leary, M. R. (2004). Reactions to acceptance and rejection: Effect of level and sequence of relational evaluation. Journal of Experimental Social Psychology, 40, 14-28

Burke, D. M., & Laver, G. D. (1990). 10 Aging and Word Retrieval: Selective Age Deficits in Language. In *Advances in psychology* (Vol. 72, pp. 281-300). North-Holland.

Burke, D. M., & Shafto, M. A. (2004). Aging and language production. *Current directions in psychological science*, *13*(1), 21-24.

Carstensen, L. L., Gottman, J. M., & Levenson, R. W. (1995). Emotional behavior in long-term marriage. *Psychology and aging*, *10*(1), 140.

Carter-Sowell, A. R., Chen, Z., & Williams, K. D. (2008). Ostracism increases social susceptibility. Social Influence, 3(3), 143-153.

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. Memory & Cognition, 24, 403– 416.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. Journal of Personality and Social Psychology, 76, 893-910.

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual review of psychology*, *64*, 285-308.

Chartrand, T. L., Maddux, W. W., & Lakin, J. L. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. *The new unconscious*, 334-361.

Cheung, E. O., Slotter, E. B., & Gardner, W. L. (2015). Are you feeling what I'm feeling? The role of facial mimicry in facilitating reconnection following social exclusion. Motivation and Emotion, 39(4), 613-630.

Chiquito, A. B. & Quesada Pacheco, M. A. (Eds.) (2014). *Actitudes lingüísticas de los hispanohablantes hacia el idioma español y sus variantes*, Bergen Language and Linguistic Studies (BeLLS), 5.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, *55*, 591-621.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American journal of mental deficiency*, *86*(2), 127-137

Clark, E. V. (1990). On the pragmatics of contrast. Journal of child language, 17(2), 417-431.

Clark, E. V. (2007). Conventionality and contrast in language and language acquisition. New directions for child and adolescent development, 2007(115), 11-23.

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. Language, 767-811.

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

Clark, H. H. (1996). *Using language*. Cambridge university press.

Clark, H. H., & Marshall, C. (1978, July). Reference diaries. In Proceedings of the 1978 workshop on Theoretical issues in natural language processing (pp. 57-63).

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A.H. Joshe, B. Webber, & I.A. Sag (Eds.), Elements of discourse understanding (pp. 10–63). Cambridge, England: Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.

Cohen, A. S., Callaway, D. A., Najolia, G. M., Larsen, J. T., & Strauss, G. P. (2012). On "risk" and reward: Investigating state anhedonia in psychometrically defined schizotypy and schizophrenia. *Journal of Abnormal Psychology*, *121*(2), 407.

De Beni, R., & Palladino, P. (2004). Decline in working memory updating through ageing: Intrusion error analyses. *Memory*, *12*(1), 75-89.

DeWall, C. N. (2013). The Oxford handbook of social exclusion. (Oxford: Oxford University Press), 9–10.

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In Advances in experimental social psychology (Vol. 33, pp. 1-40). Academic Press.

Doyle, G., Yurovsky, D., & Frank, M. C. (2016, April). A robust framework for estimating linguistic alignment in twitter conversations. In *Proceedings of the 25th international conference on world wide web* (pp. 637-648).

Dwyer, K., David, A. S., McCarthy, R., McKenna, P., & Peters, E. (2019). Linguistic alignment and theory of mind impairments in schizophrenia patients' dialogic interactions. *Psychological medicine*, 1-9.

Eckert, P., & Rickford, J. R. (Eds.). (2001). *Style and sociolinguistic variation*. Cambridge University Press.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. Science, 302(5643), 290-292.

Ettinger, U., Mohr, C., Gooding, D. C., Cohen, A. S., Rapp, A., Haenschel, C., & Park, S. (2015). Cognition and brain function in schizotypy: a selective review. *Schizophrenia bulletin*, *41*(suppl_2), S417-S426.

Fajardo, S. C. (2014). Ideas lingüísticas presentes en diccionarios latinoamericanos del siglo XIX. Aportes para la historia de la estandarización del español hablado en América. Grande López, Martín Aizpuru, & Salicio Bravo (Eds.). *Con una letra joven: Avances en el estudio de la Historiografía e Historia de la Lengua Española*. Ediciones Universidad de Salamanca, 37-44.

Folley, B. S., & Park, S. (2005). Verbal creativity and schizotypal personality in relation to prefrontal hemispheric laterality: A behavioral and near-infrared optical imaging study. *Schizophrenia research*, *80*(2-3), 271-282.

Fussell, S. R., & Krauss, R. M. (1991). Accuracy and bias in estimates of others' knowledge. European Journal of Social Psychology, 21(5), 445-454.

Gallois, C., & Callan, V. J. (1991). Interethnic accommodation: The role of norms. *Contexts of accommodation: Developments in applied sociolinguistics. Edited by Giles, H., Coupland, J. and Coupland, N., New York, NY: Cambridge University Press. 245-269.*

Gardner, W. L., Pickett, C. L., & Brewer, M. B. (2000). Social exclusion and selective memory: How the need to belong influences memory for social events. Personality and Social Psychology Bulletin, 26(4), 486-496.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in cognitive sciences*, *8*(1), 8-11.

Gerber, J., & Wheeler, L. (2009). On being rejected: A meta-analysis of experimental research on rejection. Perspectives on Psychological Science, 4(5), 468-488.

Giles, H., & Coupland, N. (1991). Language: Contexts and consequences. Thomson Brooks/Cole Publishing Co.

Giles, H., Mulac, A., Bradac, J. J., & Johnson, P. (1987). Speech accommodation theory: The first decade and beyond. *Annals of the International Communication Association*, *10*(1), 13-48.

Gregory Jr, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. J*ournal of Personality and Social Psychology*, 70(6), 1231.

Gregory, M., & Carroll, S. (2018). *Language and situation: Language varieties and their social contexts*. Routledge.

Grice, H.P. (1975). Logic and conversation. In P. Cole and J.L. Morgan (Eds.), Syntax and semantics, Vol. 3: Speech acts (pp. 225-242). New York: Seminar Press.

Gumperz, J. (1968). The speech community. International encyclopedia of the social sciences, 381–386.

Gumperz, J. J. (2009). The speech community. Linguistic anthropology: A reader, 1, 66.

Gunnar, M. R., Sebanc, A. M., Tout, K., Donzella, B., & Van Dulmen, M. M. (2003). Peer rejection, temperament, and cortisol activity in preschoolers. Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 43(4), 346-368.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. Cognitive Science, 28(1), 105-115.

Hanna, Joy E., Michael K. Tanenhaus, and John C. Trueswell. "The effects of common ground and perspective on domains of referential interpretation." Journal of Memory and Language 49, no. 1 (2003): 43-61.

Healey, M. L., & Grossman, M. (2016). Social coordination in older adulthood: A dual-process model. *Experimental aging research*, *42*(1), 112-127.

Hedden, T., & Gabrieli, J. D. (2004). Insights into the ageing mind: a view from cognitive neuroscience. *Nature reviews neuroscience*, *5*(2), 87-96.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166-1186.

Heeren, A., Dricot, L., Billieux, J., Philippot, P., Grynberg, D., de Timary, P., et al. (2017). Correlates of social exclusion in social anxiety disorder: an fMRI study. Sci. Rep. 7, 1–10. doi: 10.1038/s41598-017-00310-9

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. Cognition, 108(3), 831-836.

Hess, Y. D., & Pickett, C. L. (2010). Social rejection and self-versus other-awareness. Journal of Experimental Social Psychology, 46(2), 453-456.

Hopkins, Z., Yuill, N., & Branigan, H. P. (2017). Inhibitory control and lexical alignment in children with an autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *58*(10), 1155-1165.

Horton, W. S. (2014). Individual differences in perspective taking and field-independence mediate structural persistence in dialog. *Acta psychologica*, *150*, 41-48.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*(4), 589-606.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*(2), 127-142.

Horton, W. S., & Gerrig, R. J. (2016). Revisiting the memory based processing approach to common ground. *Topics in Cognitive Science*, 8(4), 780-795.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground?. Cognition, 59(1), 91-117.

Horton, W. S., & Spieler, D. H. (2007). Age-related differences in communication and audience design. *Psychology and Aging*, *22*(2), 281.

Jahshan, C. S., & Sergi, M. J. (2007). Theory of mind, neurocognition, and functional status in schizotypy. *Schizophrenia research*, *89*(1-3), 278-286.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, *2*(1999), 102-138.

Juncos-Rabadán, O., Facal, D., Rodríguez, M. S., & Pereiro, A. X. (2010). Lexical knowledge and lexical retrieval in ageing: Insights from a tip-of-the-tongue (TOT) study. *Language and Cognitive Processes*, *25*(10), 1301-1334.

Juncos-Rabadán, O., Pereiro, A. X., Rodríguez, N., & Facal, D. (2009). Lexical access in MCI: A study on the tip-of-the-tongue events. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *5*(4), P377.

Kaplan, O., & Lubow, R. E. (2011). Ignoring irrelevant stimuli in latent inhibition and Stroop paradigms: the effects of schizotypy and gender. *Psychiatry Research*, *186*(1), 40-45.

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, *133*(3), 450.

Kaschak, M. P., Kutta, T. J., & Jones, J. L. (2011). Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic bulletin & review*, *18*(6), 1133-1139.

Kavé, G., & Mashal, N. (2012). Age-related differences in word-retrieval but not in meaning generation. *Aging, Neuropsychology, and Cognition*, *19*(4), 515-529.

Kemp, K. C., Gross, G. M., & Kwapil, T. R. (2019). Psychometric Properties of the Multidimensional Schizotypy Scale and Multidimensional Schizotypy Scale–Brief: Item and Scale Test–Retest Reliability and Concordance of Original and Brief Forms. *Journal of personality assessment*, 1-8.

Kemper, S., & Anagnopoulos, C. (1989). Language and aging. *Annual review of applied linguistics*, *10*, 37-50.

Keysar, Boaz, 2007. Communication and miscommunication: the role of egocentric processes. Intercultural Pragmatics 4 (1), 71–84.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. Psychological Science, 11(1), 32-38.

Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. Journal of Memory and Language, 39(1), 1-20.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. Cognition, 89(1), 25-41.

Kiang, M., & Kutas, M. (2005). Association of schizotypy with semantic processing differences: an event-related brain potential study. *Schizophrenia Research*, *77*(2-3), 329-342.

Kiang, M., & Kutas, M. (2006). Abnormal typicality of responses on a category fluency task in schizotypy. *Psychiatry Research*, *145*(2-3), 119-126.

Kiang, M., Prugh, J., & Kutas, M. (2010). An event-related brain potential study of schizotypal personality and associative semantic processing. *International Journal of Psychophysiology*, *75*(2), 119-126.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in cognitive sciences*, *22*(2), 154-169.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

Krauss, R. M., & Glucksberg, S. (1969). The development of communication: Competence as a function of age. *Child Development*, 40, 255-266.

Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language*, 83, 1-19.Kumari V, Ettinger U. Latent inhibition in schizophrenia and schizotypy: a review of the empirical literature. In: Lubow RE, ed. Latent Inhibition. Cambridge, UK: Cambridge University Press; 2010:419–447

Kumari, V., & Ettinger, U. (2010). Latent inhibition in schizophrenia and schizotypy: a review of the empirical literature. In Lubow, R.E. (Ed) *Latent inhibition*, Cambridge, UK: Cambridge University Press;419-447.

Kuperberg, G. R. (2010a). Language in schizophrenia part 1: an introduction. *Language and linguistics compass*, *4*(8), 576-589.

Kuperberg, G. R. (2010b). Language in schizophrenia Part 2: What can psycholinguistics bring to the study of schizophrenia… and vice versa? *Language and linguistics compass*, *4*(8), 590-604.

Kwapil, T. R., & Barrantes-Vidal, N. (2015). Schizotypy: looking back and moving forward. *Schizophrenia bulletin*, *41*(suppl_2), S366-S373.

Kwapil, T. R., Brown, L. H., Silvia, P. J., Myin-Germeys, I., & Barrantes-Vidal, N. (2012). The expression of positive and negative schizotypy in daily life: an experience sampling study. *Psychological Medicine*, *42*(12), 2555-2566.

Kwapil, T. R., Gross, G. M., Silvia, P. J., Raulin, M. L., & Barrantes-Vidal, N. (2018). Development and psychometric properties of the Multidimensional

Schizotypy Scale: A new measure for assessing positive, negative, and disorganized schizotypy. *Schizophrenia research*, *193*, 209-217.

Labov, W. (1972). Sociolinguistic Patterns. Philadelphia: University of Pennsylvania Press.

Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. Psychological Science, 14(4), 334–339. https://doi.org/10.1111/1467-9280.14481

Lakin, J. L., & Chartrand, T. L. (2013). Behavioral mimicry as an affiliative response to social exclusion. The Oxford handbook of social exclusion, 266-274.

Lakin, J. L., Chartrand, T. L., & Arkin, R. M. (2008). I am too just like you: Nonconscious mimicry as an automatic behavioral response to social exclusion. *Psychological science*, *19*(8), 816-822.

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. Journal of Nonverbal Behavior, 27(3), 145–162.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174.

Langdon, R., & Coltheart, M. (2004). Recognition of metaphor and irony in young adults: the impact of schizotypal personality traits. *Psychiatry research*, *125*(1), 9-20.

Larson, M. J., Baldwin, S. A., Good, D. A., & Fair, J. E. (2010). Temporal stability of the error related negativity (ERN) and post error positivity (Pe): The role of number of trials. *Psychophysiology*, *47*(6), 1167-1171.

Leaper, C., & Ayres, M. M. (2007). A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. *Personality and Social Psychology Review*, *11*(4), 328-363.

Leighton, J., Bird, G., Orsini, C., & Heyes, C. (2010). Social attitudes modulate automatic imitation. Journal of Experimental Social Psychology, 46(6), 905-910.

Lev-Ari, S. (2018). The influence of social network size on speech perception. *Quarterly Journal of Experimental Psychology*, 71(10), 2249-2260.

Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations.

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard, University Press.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. Psychonomic bulletin & review, 9(3), 550-557.

Logie, R. H., & Morris, R. G. (Eds.). (2014). *Working memory and ageing*. Hove, UK: Psychology Press.

Long, M. R., Horton, W. S., Rohde, H., & Sorace, A. (2018). Individual differences in switching and inhibition predict perspective-taking across the lifespan. *Cognition*, *170*, 25-30.

Lustig, C., Hasher, L., & Tonev, S. T. (2001). Inhibitory control over the present and the past. *European Journal of Cognitive Psychology*, *13*(1-2), 107-122.

Lynn, R. & Martin, T. (1997) Gender Differences in Extraversion, Neuroticism, and Psychoticism in 37 Nations, The Journal of Social Psychology, 137:3, 369-373, DOI: 10.1080/00224549709595447

Lysander, K., & Horton, W. S. (2012). Conversational grounding in younger and older adults: The effect of partner visibility and referent abstractness in task-oriented dialogue. *Discourse Processes*, *49*(1), 29-60.

MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological bulletin*, *131*(2), 202.

Mallott, M. A., Maner, J. K., DeWall, N., and Schmidt, N. B. (2009). Compensatory deficits following rejection: the role of social anxiety in disrupting affiliative behaviour. Depress. Anxiety 26, 438–446. doi: 10.1002/da.20555

Maner, J. K., DeWall, C. N., Baumeister, R. F., & Schaller, M. (2007). Does social exclusion motivate interpersonal reconnection? Resolving the" porcupine problem.". Journal of personality and social psychology, 92(1), 42.

March, E., Wales, R., & Pattison, P. (2003). Language use in normal ageing and dementia of the Alzheimer type. *Clinical Psychologist*, 7(1), 44-49.

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. Behavior research methods, 43(3), 679-690.

Matsumoto, Y., Yamagishi, T., Li, Y., & Kiyonari, T. (2016). Prosocial behavior increases with age across five economic games. *PloS one*, *11*(7).

Mauersberger, H., Blaison, C., Kafetsios, K., Kessler, C. L., & Hess, U. (2015). Individual differences in emotional mimicry: Underlying traits and social consequences. *European Journal of Personality*, *29*(5), 512-529.

Mayr, U., Spieler, D. H., & Kliegl, R. (Eds.). (2001). *Ageing and executive control* (Vol. 13, No. 1-2). Psychology Press.

McDonald, M. M., & Donnellan, M. B. (2012). Is ostracism a strong situation? The influence of personality in reactions to rejection. Journal of Research in Personality, 46(5), 614-618.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, *1*(1), 30.

Mead, N. L., Baumeister, R. F., Stillman, T. F., Rawn, C. D., & Vohs, K. D. (2011). Social exclusion causes people to spend and consume strategically in the service of affiliation. Journal of consumer research, 37(5), 902-919.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49(2), 201-213.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture–word interference experiments. *Journal of memory and Language*, *35*(4), 477-496.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology*, *90*(2), 227.

Mitchell, D. B. (1989). How many memory systems? Evidence from aging. Journal of Experimental Psychology; Learning, Memory, & Cognition, 15, 31-49.

Monson, T. C., Hesley, J. W., & Chernick, L. (1982). Specifying when personality traits can and cannot predict behavior: An alternative to abandoning the attempt to predict single-act criteria. Journal of Personality and Social Psychology, 43(2), 385.

Morett, Sonia (2014). Actitudes lingüísticas en México. Entre el chovinismo y el malinchismo. Chiquito, A. B. & Quesada Pacheco, M. A. (eds.). *Actitudes lingüísticas de los hispanohablantes hacia el idioma español y sus variantes*, Bergen Language and Linguistic Studies (BeLLS), 5. Available at http://dx.doi.org/10.15845/bells.v5i0.688

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. Psychological Science, 13(4), 329-336.

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non Gaussian data: a practical guide for biologists. *Biological Reviews*, *85*(4), 935-956.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. Journal of Personality and Social Psychology, 32(5), 790.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, *4*(5), 648-654.

Nettle, D. (1999). Using social impact theory to simulate language change. *Lingua*, 108(2-3), 95-117.

Newby, J., Pitura, V. A., Penney, A. M., Klein, R. G., Flett, G. L., & Hewitt, P. L. (2017). Neuroticism and perfectionism as predictors of social anxiety. *Personality and Individual Differences*, *106*, 263-267.

Nicholas, M., Obler, L., Albert, M., & Goodglass, H. (1985). Lexical retrieval in healthy aging. *Cortex*, *21*(4), 595-606.

Nickerson, R. S., Baddeley, A., & Freeman, B. (1987). Are people's estimates of what other people know influenced by what they themselves know?. Acta Psychologica, 64(3), 245-259.

Oaten, M., Williams, K. D., Jones, A., and Zadro, L. (2008). The effects of ostracism on self-regulation in the socially anxious. J. Soc. Clin. Psychol. 27, 471–504. doi: 10.1521/jscp.2008.27.5.471

Palomares, N. A., Giles, H., Soliz, J., & Gallois, C. (2016). Intergroup accommodation, social categories, and identities. *Giles, H. (Ed.), Communication accommodation theory: Negotiating personal relationships and social identities across contexts, Cambridge University Press,* 123-151.

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, *79*(2), 637-659.

Park, S., Lenzenweger, M. F., Püschel, J., & Holzman, P. S. (1996). Attentional inhibition in schizophrenia and schizotypy: A spatial negative priming study. *Cognitive Neuropsychiatry*, *1*(2), 125-150.

Peters, E. R., Pickering, A. D., & Hemsley, D. R. (1994). ‘Cognitive inhibition’and positive symptomatology in schizotypy. *British Journal of Clinical Psychology*, *33*(1), 33-48.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and language*, *39*(4), 633-651.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169-190.

Polit, D. F. (2015). Assessing measurement in health: beyond reliability and validity. *International journal of nursing studies*, *52*(11), 1746-1753.

Prévost, M., Rodier, M., Renoult, L., Kwann, Y., Dionne, Dostie, E., Chapleau, I., ... & Debruille, J. B. (2010). Schizotypal traits and N400 in healthy subjects. *Psychophysiology*, *47*(6), 1047-1056.

Rabagliati, H., Delaney-Busch, N., Snedeker, J., & Kuperberg, G. (2019). Spared bottom-up but impaired top-down interactive effects during naturalistic language processing in schizophrenia: evidence from the visual-world paradigm. *Psychological medicine*, *49*(8), 1335-1345.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111-163.

Rapp, A. M., Langohr, K., Mutschler, D. E., & Wild, B. (2014). Irony and proverb comprehension in schizophrenia: do female patients “dislike” ironic remarks?. *Schizophrenia research and treatment*. Available online: http://dx.doi.org/10.1155/2014/841086.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, 85(2), 59.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. Psychological review, 95(3), 385.

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. Journal of Memory and Language, 76, 29-46.

Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition*, *125*(1), 49-63.

Schiffer, S. (1972). *Meaning*. Oxford: Clarendon Press.

Schmidt-Hansen, M., & Honey, R. C. (2009). Working memory and multidimensional schizotypy: dissociable influences of the different dimensions. *Cognitive neuropsychology*, *26*(7), 655-670.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609-612.

Segaert, K., Kempen, G., Petersson, K. M., & Hagoort, P. (2013). Syntactic priming and the lexical boost effect during sentence production and sentence comprehension: An fMRI study. *Brain and language*, *124*(2), 174-183.

Shao, Z., Roelofs, A., & Meyer, A. S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. The Quarterly Journal of Experimental Psychology, 65(10), 1927-1944.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of memory and language*, *81*, 105-120.

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. Behavior research methods, 49(2), 418-432.

Stel, M., & Vonk, R. (2010). Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction. British Journal of Psychology, 101(2), 311-323.

Strauss, G. P., Allen, D. N., Jorgensen, M. L., & Cramer, S. L. (2005). Test-retest reliability of standard and emotional stroop tasks: an investigation of color-word and picture-word versions. *Assessment*, *12*(3), 330-337.

Swickert, R., & Owens, T. (2010). The interaction between neuroticism and gender influences the perceived availability of social support. Personality and Individual Differences, 48(4), 385-390.

Tandon, R., Nasrallah, H. A., & Keshavan, M. S. (2009). Schizophrenia,"just the facts" 4. Clinical features and conceptualization. *Schizophrenia research*, *110*(1-3), 1-23.

Team, R. C. (2013). R: A language and environment for statistical computing.

Tennant, C., Bebbington, P., & Hurry, J. (1982). Female vulnerability to neurosis: the influence of social roles. Australian and New Zealand Journal of Psychiatry, 16(3), 135-140.

Thakerar, J. N., Giles, H., & Cheshire, J. (1982). Psychological and linguistic parameters of speech accommodation theory. In Fraser, C. & Scherer, K. R. (Eds), *Advances in the Social Psychology of Language*. UK: Cambridge University Press. 205-255.

Thomas, J.C., Fozard, J.L., & Waugh, N.C. (1977). Age-related differences in naming latency. *American Journal of Psychology*, 90, 499-509.

Tobar-Henríquez, A., Rabagliati, H., & Branigan, H. P. (2019). Lexical entrainment reflects a stable individual trait: Implications for individual differences in

language processing. *Journal of experimental psychology. Learning, memory, and cognition.* Advance online publication: https://doi.org/10.1037/xlm0000774

Tonelli, H. A. (2014). How semantic deficits in schizotypy help understand language and thought disorders in schizophrenia: a systematic and integrative review. *Trends in psychiatry and psychotherapy*, *36*(2), 75-88.

Traxler, M. J., Tooley, K. M., & Pickering, M. J. (2014). Syntactic priming during sentence comprehension: Evidence for the lexical boost. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 905.

Twenge, J. M., & Campbell, W. K. (2003). "Isn't it fun to get the respect that we're going to deserve?" Narcissism, social rejection, and aggression. Personality and Social Psychology Bulletin, 29(2), 261-272.

Van Baaren, R. B., Holland, R. W., Steenaert, B., & van Knippenberg, A. (2003). Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology*, *39*(4), 393-398.

Vollema, M. G., & van den Bosch, R. J. (1995). The multidimensionality of schizotypy. *Schizophrenia bulletin*, *21*(1), 19-31.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic Bulletin & Review*, *14*(5), 779-804.

Weerdmeester, J., & Lange, W. G. (2019). Social anxiety and pro-social behavior following varying degrees of rejection: Piloting a new experimental paradigm. *Frontiers in psychology*, 10:1325. doi: 10.3389/fpsyg.2019.01325

Williams, K. D. (2007). Ostracism: The kiss of social death. *Social and Personality Psychology Compass*, *1*(1), 236-247.

Williams, K. D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. Behavior research methods, 38(1), 174-180.

Williams, K. D., & Sommer, K. L. (1997). Social ostracism by coworkers: Does rejection lead to loafing or compensation?. Personality and Social Psychology Bulletin, 23(7), 693-706.

Williams, K. D., Cheung, C. K., & Choi, W. (2000). Cyberostracism: effects of being ignored over the Internet. *Journal of personality and social psychology*, *79*(5), 748.

Williams, K. D., Govan, C. L., Croker, V., Tynan, D., Cruickshank, M., & Lam, A. (2002). Investigations into differences between social- and cyberostracism. Group Dynamics: Theory, Research, and Practice, 6(1), 65–77.

Xu, Y., & Reitter, D. (2015, June). An evaluation and comparison of linguistic alignment measures. In *Proceedings of the 6th Workshop on Cognitive Modeling and Computational Linguistics* (pp. 58-67).

Zadro, L., Williams, K. D., & Richardson, R. (2004). How low can you go? Ostracism by a computer is sufficient to lower self-reported levels of belonging, control, self-esteem, and meaningful existence. Journal of Experimental Social Psychology, 40(4), 560–567. https://doi.org/10.1016/j.jesp.2003.11.006