

Speakers extrapolate community-level knowledge from individual linguistic encounters.

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, and by ESRC grants ES/N005635/1 and ES/N013115/1

All analyses, data, and stimuli are available at <https://github.com/anitatobar/lingcommunities>.

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

Abstract

Speakers vary their lexical choices based on individual-level influences, for example, tending to reuse the same words as their interlocutors. However, it is unclear how lexical choices are affected by community-level factors, e.g., whether the interlocutor is from their own speech community (*in-community partner*) or from another community (*out-community partner*). Indeed, we know very little about how speakers establish community-level linguistic knowledge in general. In three experiments, we examined (i) how speakers' referential choices varied depending on their partner's choices and speech community, and (ii) how speakers' extrapolation of their own choices to a subsequent partner was modulated by their partners' speech communities. In Experiment 1, Spanish participants played two sessions of an online picture-matching-and-naming task, using the same pictures but with a different confederate in each session. In Session 1, the confederate was either an *in-community partner* (Spanish) or an *out-community partner* (Latin American); in Session 2, the confederate was either from the same community as the first partner or not. In Session 1, participants' referential choices were influenced by the confederate's choices, but not by their community. In Session 2, however, participants generalised their expressions from Session 1 based on their confederates' communities. Experiment 2 replicated this pattern of results in Mexican participants, and Experiment 3 confirmed that these results were driven by differences in confederates' communities, rather than perceived linguistic status. Taken together, these results suggest that speakers encode speech community information during language processing and store that information to inform future contexts of language use, even when such community information has not affected speakers' language use during that particular linguistic encounter. Critically, these results show that speakers establish community-level knowledge by extrapolating linguistic information from individual-level experiences.

Dialogue, language production, lexical entrainment, alignment, speech community, common ground

Speakers extrapolate community-level knowledge from individual linguistic encounters.

Language use is fundamentally variable: The same object can be called *potato* or *spud* in English, or *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 is likely to be more fragmented; for instance, they may not realise that, in Mexico, *papa* is the favoured term. Investigating how speakers use language with conversational partners from other speech communities offers an opportunity to investigate how individual- and community-level influences interact, and moreover how people extrapolate from individual-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 individual- 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 entrainment may also imply an audience design component. Under the assumption that different lexical labels reflect different conceptualisations (see E. V., Clark, 1987), a speaker's reuse of a partner's term can signal that they agree with their partner's proposed conceptualisation of a referent (Brennan & Clark, 1996, p. 1491-1492). Under this logic, when speakers reuse a partner's term, they do so in order to accommodate their use of language to their partner's expectations, thus facilitating mutual comprehension. Importantly, audience design accounts of entrainment 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). But such accounts do argue that speakers' tendency to reuse their partner's terms is dependent on their beliefs about their partner's expectations of language use.

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 individual interactions with in-community (Castilian) partners, Castilian children learn Castilian community preferences, developing in-community-level knowledge; e.g., that *patata* is the Castilian favoured term for potato, though some individual

Castilian speakers may prefer *papa* (Clark, 1996; Clark, 2007). Interestingly, previous work suggests that the more input speakers have had from different partners, the less sensitive their 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 (e.g., that a friend's use of *papa* is an idiosyncratic preference that is at odds with their community's preference for *patata*).

These considerations may constrain lexical entrainment. If an in-community speaker names an object with a disfavoured term (e.g., *papa*) instead of a favoured term (i.e., *patata*) when interacting with an in-community addressee, then that addressee will know that this usage is an idiosyncratic preference of that particular speaker, which does not match the community's preference. Since members in the same speech community would be expected to be familiar with (hence, understand) their own community's favoured terms, a Castilian speaker's use of a disfavoured term like *papa* would not necessarily motivate a Castilian partner to reuse that term to enhance communication. Moreover, experiencing an in-community speaker using a disfavoured 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 disfavoured term with a subsequent in-community speaker, unless its use was promoted by transient low-level priming effects.

Similarly, through individual linguistic interactions with members of other communities such as Mexicans (i.e., *out-community partners*), Castilians can establish knowledge about language use in those communities. 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, individuals will have weaker knowledge about whether an out-community partner's term is an individual or a community preference. For example,

if a Mexican used a disfavoured term (e.g., *papa*), a Castilian's knowledge of Mexican community preferences might not be accurate enough for them to judge accurately whether *papa* is an idiosyncratic use or an out-community-level (Mexican) preference (as is in fact the case, so that *papa* is the favoured term for Mexican speakers). Given this uncertainty, an out-community partner's use of a disfavoured term may increase speakers' likelihood to reuse that term (i.e., entrain) with the same partner for the purpose of ensuring mutual comprehension.

Importantly, given their limited experience with out-community partners' community-level preferences, speakers may be sensitive to new input from an out-community partner (Lev-Ari, 2018), so that a single linguistic interaction with a Mexican may meaningfully update a Castilian's knowledge of Mexicans' community-level preferences, for instance strengthening the belief that Mexicans have a community preference for the (Castilian disfavoured) term *papa* over the (Castilian favoured) term *patata*. This will in turn increase the Castilian's likelihood to use that disfavoured term *papa* with subsequent Mexican partners (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 can be 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 favoured term (e.g., *potato*) and a disfavoured (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 disfavoured 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 favoured 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 disfavoured term as representative of a computer-community preference, and then assume that computers might not understand the favoured term, increasing the likelihood of lexical entrainment on the disfavoured term. In contrast, participants' extensive previous experience with native English-speaking humans would mean that they would interpret a human's use of a disfavoured term as an idiosyncratic preference, and would not assume that their human-partner would not understand the favoured 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.

Nevertheless, 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 that referent than when they have not (e.g., Metzing & Brennan, 2003; see Krönmüller & 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 information about a partner's community is encoded along with lexical information during dialogue. More specifically, it is unclear whether speakers encode partners' community information during individual linguistic experiences in ways that immediately affect their referential expressions, and whether they extrapolate from these individual experiences to inferences about likely community preferences that might affect their referential expressions with future partners.

The present experiments

To address these issues, we investigated the effects of participants' beliefs about their partner's speech community on their referential expressions, as manifested in 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 favoured term (e.g., *patata*) and a disfavoured but acceptable term (e.g., *papa*) (as established by a pretest).

In the first session, the partner always named the target first, using the disfavoured 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 disfavoured 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 disfavoured 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 disfavoured 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 disfavoured 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. Such a pattern would provide

evidence that speakers encode information about community membership during lexical processing in dialogue.

We similarly expected that participants would tend to maintain entrained terms, so that they would be more likely to use the disfavoured term following previous use than in a spontaneous naming task. But we expected that the extent to which participants maintained entrained terms would vary between conditions. If speakers are more likely to extrapolate a (disfavoured) entrained term from one partner to another partner from that community when they are less confident about community preferences, then participants should maintain a precedent more often when interacting with an out-community partner after entraining to an out-community partner than when interacting with an in-community partner. However, they should not maintain a precedent more often when interacting with an in-community partner after entraining to an in-community partner than when interacting with an out-community partner. Such a pattern would provide evidence that speakers encode information about community membership during dialogue, and moreover that they use such information to establish community-level knowledge from individual experiences.

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 disfavoured term that their partner had used earlier (Session 1; Lexical Entrainment), and whether they reused disfavoured 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 generalise (i.e., maintain) disfavoured 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 generalise (i.e., maintain) disfavoured terms from an in-community partner to another in-community partner more often than to an out-community partner.

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 disfavoured terms in Session 1) and (b) Maintenance of Entrained Terms (i.e., participants' reuse during Session 2 of disfavoured terms that they had used in Session 1).

Materials. In both sessions, we used 21 target pictures that could be labelled with either a favoured term or a disfavoured 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. 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 favoured 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 disfavoured 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 favoured term, and at least 10% of participants had provided the same disfavoured 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 disfavoured terms with respect to the pictures, on a scale from 1 to 7. We used this to create the final set of 21 disfavoured terms (see Appendix), each of which had been spontaneously used with a frequency below 30% when answering the favoured 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 favoured term task to choose 21 filler pictures, in which 80% of participants agreed on the same favoured term.

Procedure. Participants completed two sessions of an online matching-and-naming task. On each trial, they were shown two pictures (see Figure 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 favoured term (e.g., *patata* [potato]) or a disfavoured, but still acceptable, term (e.g., *papa*).

The structure of the matching and naming task is illustrated in Figure 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 disfavoured 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.

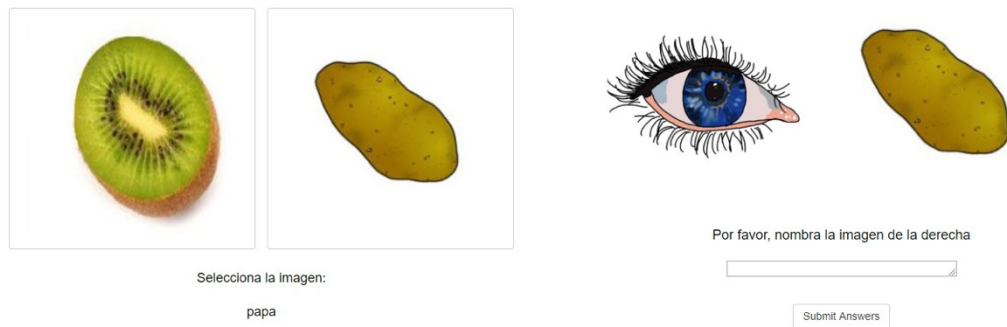
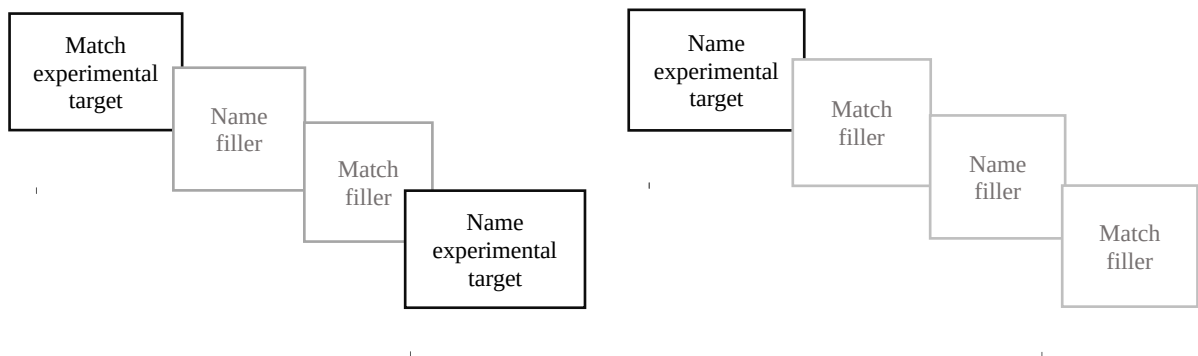
A**B**

Figure 1. A. Examples of matching and naming trials (where the favoured term is *patata* and disfavoured 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 disfavoured 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 interacted with different partners across sessions.

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 1).

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.

Results

Data processing and exclusions. In Session 1, we coded all naming trials for whether they showed Lexical Entrainment (using the disfavoured 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). 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 given data 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 disfavoured 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 disfavoured terms in 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 disfavoured 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 disfavoured terms in Session 1 (i.e., lexical entrainment trials). In particular, we tested whether maintenance of disfavoured terms was affected by participants' beliefs about their partner's speech community, by regressing the use of disfavoured 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 disfavoured term on approximately half of the trials (50%[SD=30%] by-participants and 52%[17%] by-items). The by-items Wilcoxon test indicated that disfavoured 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 2).

Critically, the degree to which participants lexically entrained was not significantly affected by their First Partner's Community (see Table 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 was not be affected by a partner's speech community (see Figure 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 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 $e(\text{BIC}_{\text{alternative}} - \text{BIC}_{\text{null}})/2$ (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 ($\text{BF} / (\text{BF} + 1) = .98$), 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 1). We found no evidence for such differences: Participants entrained to their first partner at similar rates when their second partner was go-

ing 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.

Overall, these results suggest that speakers entrain to their conversational partners’ terms even during remote computer-mediated interactions, but show no evidence that this tendency is mediated by the speech communities of those partners.

Table 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

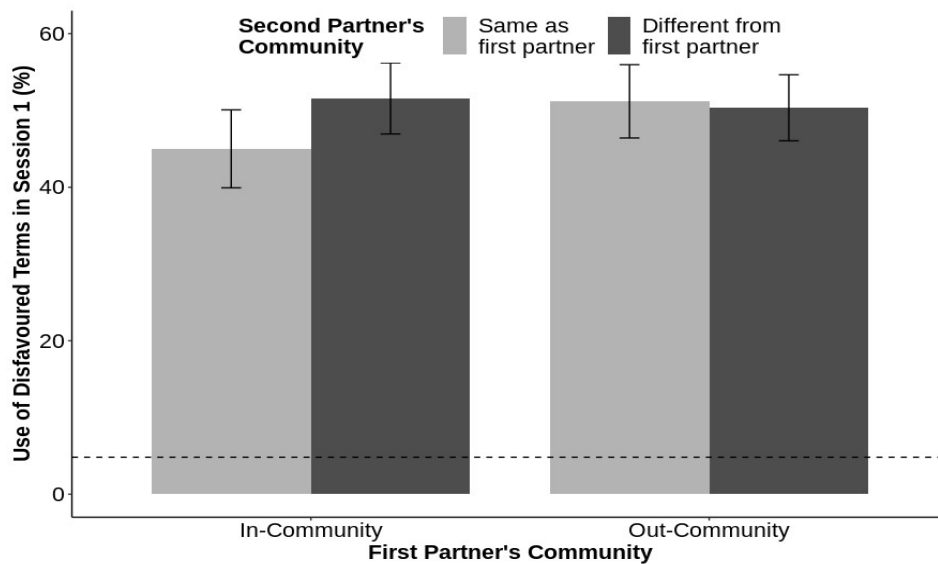


Figure 2. Mean and standard error of the percentage of use of disfavoured terms in Session 1 (y-axis) across First Partner’s Community (x-axis) and Second Partner’s Community (color-coded). The horizontal dashed line represents the mean of percentage of use of disfavoured terms on the pretest.

Maintenance of Entrained Terms. Overall, participants reused disfavoured 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) disfavoured terms in Session 2 than participants who named the objects spontaneously during the norming task (4%[6%]; $V=0$, $p<.0001$; see Figure 3). However, they did not maintain terms perfectly: Participants were significantly less likely to reuse disfavoured terms in Session 2 than to initially use such terms in Session 1 ($V=0$, $p<.0001$), and for items on which the disfavoured term was used in Session 1, participants maintained that disfavoured term on 70%[27%] of trials in Session 2. Figure 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 2): Participants maintained disfavoured terms at similar rates across two partners from different communities (68%[29%]) as across two partners from the same community (72%[26%]).

There was a significant effect of First Partner's Community ($\beta=-.36$, $SE=.13$, $z=-2.7$, $p<.01$, see Table 2): Participants maintained disfavoured terms more often in Session 2 after playing with an in-community partner in Session 1 (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 ($\beta=-.3$, $SE=.13$, $z=-2.3$, $p=.02$, see Table 2 and Figure 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 2. *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

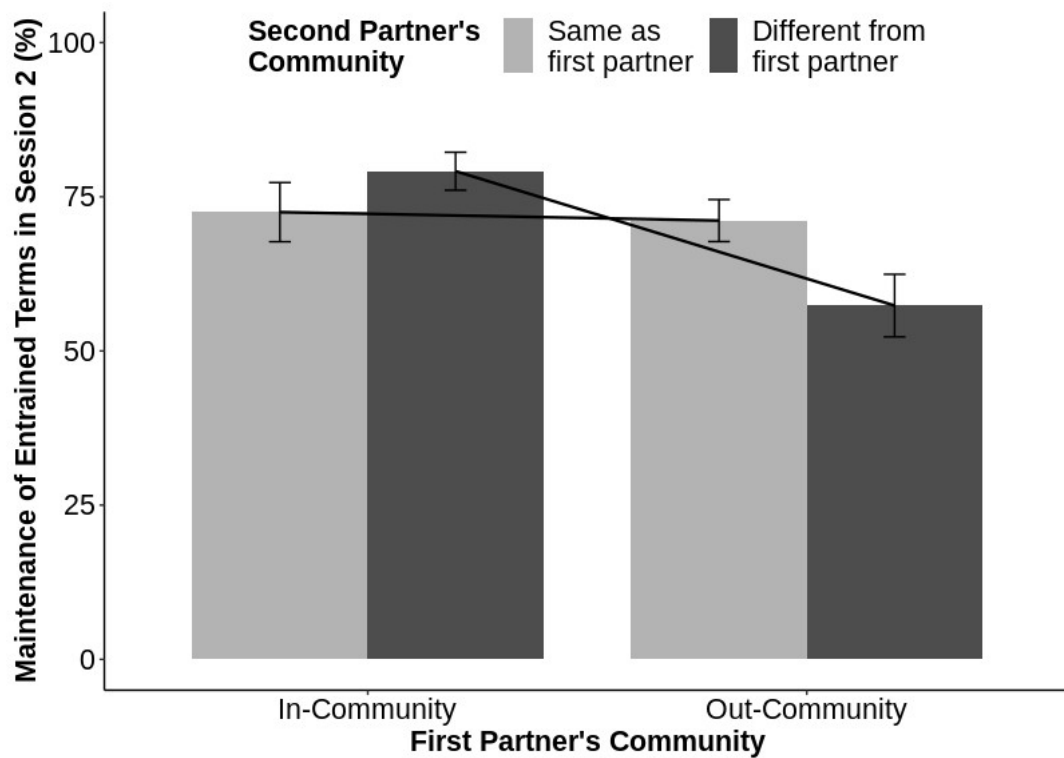


Figure 3. Mean and standard error of percentage of (re)use of disfavoured terms in Session 2 (y-axis) across First Partner's Community (x-axis) and Second Partner's Community (color-coded). Solid lines represent a significant interaction between the First Partner's Community and Second Partner's Community 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 3 and Figure 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 4 and Figure 3).

Table 3. 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 4. 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.

Discussion

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 disfavoured 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 behaviour even when it does not affect their concurrent behaviour. Specifically, participants extrapolated referential expressions that had been used by out-community 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 partners from their own community than to out-community partners. 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 especially 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' favoured terms, then we would expect both that Mexicans would entrain more to (Argentinian) out-community partners than to (Mexican) in-community partners, and that they would not extrapolate (Mexican) in-community partners' expressions to (Argentinian) out-community partners at similar rates as to other (Mexican) in-community partners. But if they were due to a general tendency for speakers to entrain irrespective of a partner's speech community, but to extrapolate preferences with regard to speech community, then Mexicans should entrain to (Argentinian) out-community partners to the same extent as to (Mexican) in-community partners, and should extrapolate (Mexican) in-community partners' expressions to (Argentinian) out-community partners at similar rates as to other (Mexican) in-community partners.

Experiment 2: Do the effects of partner's community membership on speakers' lexical expressions generalise 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.

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 favoured and disfavoured terms, and 100 new Mexican individuals (aged 18-54, $M=31$, $SD=7$) rated acceptability of the disfavoured terms. We

created a final set of 20 disfavoured terms (see Appendix); mean use as a favoured 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 favoured 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 favoured and a disfavoured term. The structure of the matching and naming task was the same as in Experiment 1 (Figure 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).

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 disfavoured 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 disfavoured terms was significantly higher than in the norming task (4%[6%]; $V=0$, $p<.0001$) (see Figure 4).

As in Experiment 1, the degree to which participants lexically entrained was not significantly affected by First Partner's Community (see Table 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%]). 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 evidence 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 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.

Overall, these results confirm Experiment 1's findings that speakers entrain to their partners' terms even during remote computer-mediated interactions, but again show no evidence that this tendency is mediated by those partners' speech communities.

Table 5. 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

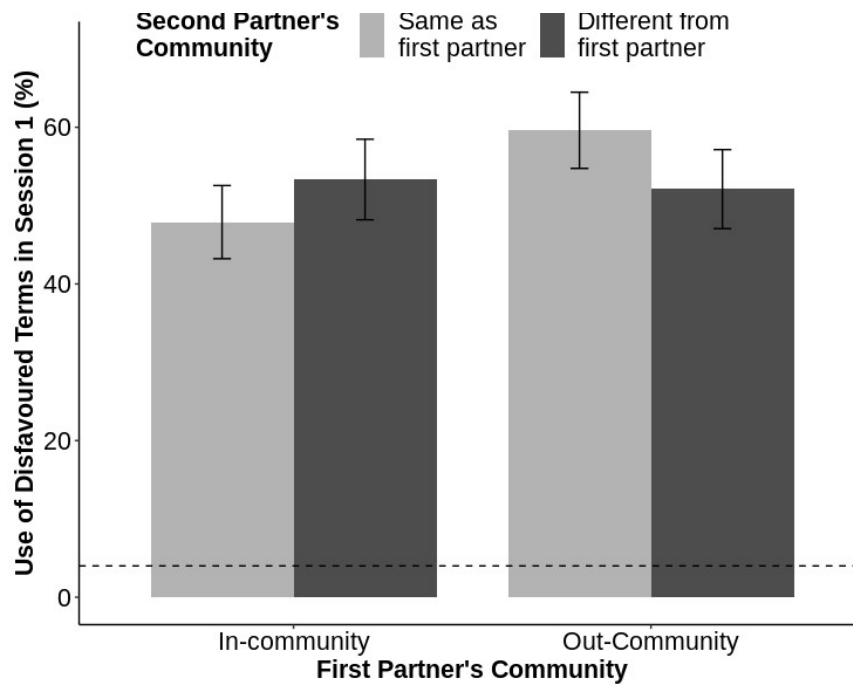


Figure 4. Mean and standard error of the percentage of use of disfavoured terms in Session 1 (y-axis) across First Partner's Community (x-axis) and Second Partner's Community (color-coded). The horizontal dashed line represents the mean of percentage of use of disfavoured terms on the pretest.

Maintenance of Entrained Terms. As in Experiment 1, in Session 2 participants used disfavoured 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) disfavoured 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 disfavoured term in Session 2 than to initially use them in Session 1 ($V=0$, $p<.0001$), and for the items on which participants used the disfavoured term in Session 1, they maintained disfavoured terms on 68%[28%] of trials. Figure 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 ($\beta=.004$, $SE=.14$, $z=0.03$, $p>.05$, see Table 4): Participants maintained disfavoured 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 ($\beta=-.01$, $SE=.14$, $z=-.1$, $p>.05$, see Table 5): Participants maintained disfavoured terms at similar rates after playing with an in-community partner in Session 1 ($M=69\%$, $SD=28\%$) and with 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 ($\beta=-.3$, $SE=.13$, $z=-2.3$, $p=.02$, see Table 4). 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.

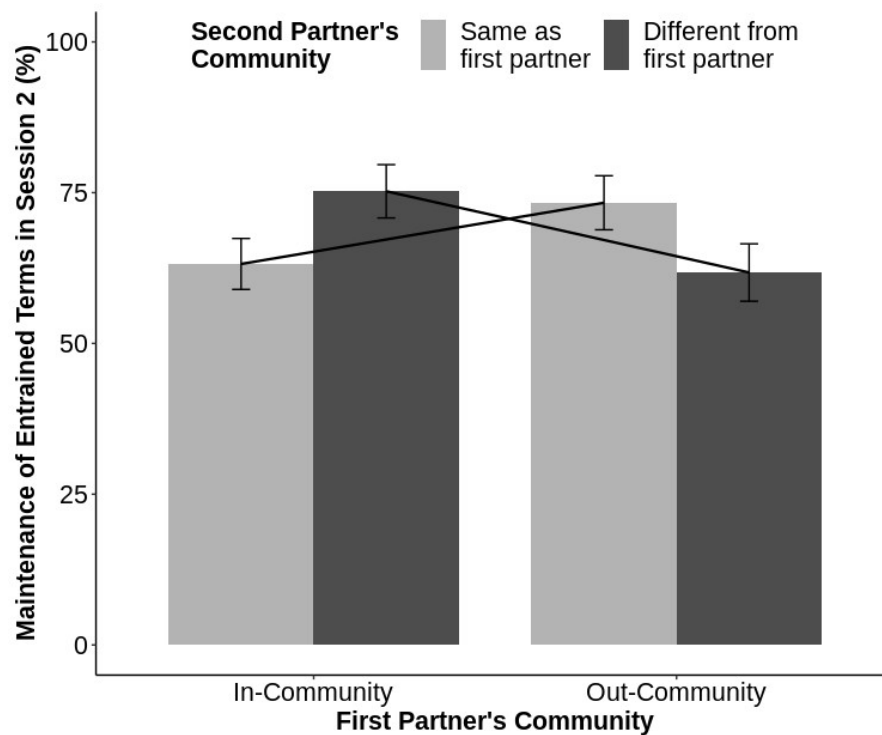


Figure 5. Mean and standard error of percentage of (re)use of disfavoured terms in Session 2 (y-axis) across First Partner's Community (x-axis) and Second Partner's Community (color-coded). Solid lines represent a significant interaction between the First Partner's Community and Second Partner's Community on maintenance of entrained terms.

Table 5. *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%]; $\beta=0.45$, $SE=.18$, $z=2.5$, $p=.012$, see Table 6 and Figure 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%]; $\beta=-.49$, $SE=.21$, $z=-2.3$, $p=.021$, see Table 7 and Figure 4).

Table 6. *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 7. *Out-Community Partner Trials: Maintenance ~ SecondPartner + (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.

Discussion

Experiment 2 replicated Experiment 1 both in finding no evidence that speakers' beliefs about a partner's speech community affect their likelihood of entraining on the same term, and in suggesting that these beliefs nevertheless 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 the possibility that Experiment 1's results 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 manipulation of in-community versus out-community partner correlated with a hierarchy of 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 that Mexicans (Experiment 2 in-community partners) perceive Argentinian Spanish (spoken by Experiment 2 out-community partners) more negatively than Mexican 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 potential effects of beliefs about speech community membership are in fact confounded with effects 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 in-community partners.

Experiment 3

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 disfavoured terms introduced by high-status Castilian partners than by lower-status Argentinian partners.

Method

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

Participants. This experiment used the responses of 40 participants from Experiment 2, who had interacted with an Argentinian out-community partner in Session 1 and a Mexican in-community partner in Session 2. We recruited 40 additional 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 disfavoured 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 1B). In the first session, we manipulated participants' beliefs about whether their partner was a member of a lower-

status speech community (i.e., Argentina) 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.

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 disfavoured term on approximately half of the trials (55%[32%] by-participants and 55%[11%] by-items). The frequency of use of the disfavoured terms was significantly higher than in the norming task (4%[6%]; $V=0$, $p<.0001$) (see Figure 6).

However, we found no significant effect of Partner’s Community on Lexical Entrainment ($\beta=.09$, $SE=.24$, $z=.38$, $p>.05$, see Table 5). Participants were as likely to entrain to a partner’s use of a disfavoured 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 5. Session 1 Trials: LexicalEntrainment \sim 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

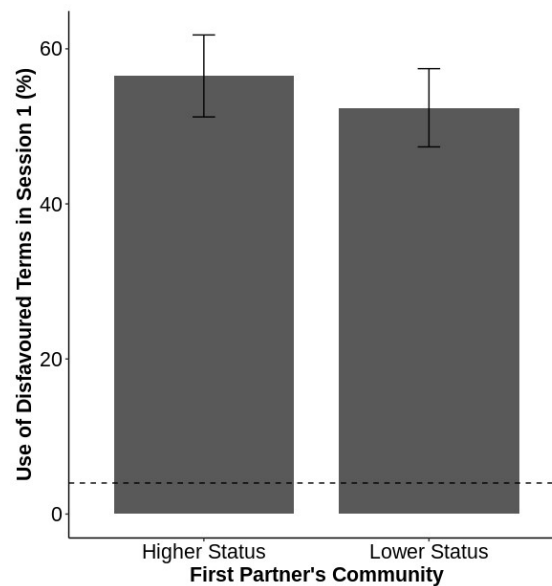


Figure 6. Mean and standard error of the percentage of use of disfavoured terms in Session 1 (y-axis) across First Partner's Community (x-axis). The horizontal dashed line represents the mean of percentage of use of disfavoured terms on the pretest.

Maintenance of Entrained Terms. As in Experiments 1 and 2, in Session 2 participants used disfavoured terms on approximately one third of trials (28%[10%] of naming trials). Importantly, participants tended to maintain entrained terms: Participants reused (entrained) disfavoured 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 disfavoured term in Session 1, participants maintained that disfavoured term 68%[26%] of the time.

However, there was no significant effect of First Partner's Community ($\beta=.24$, $SE=.23$, $z=1.02$, $p>.05$, see Table 6 and Figure 7). Participants maintained a disfavoured term they had previously used with a partner from a lower-status community in Session 1 (62%[29%]) as often as a disfavoured 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 linguistic status on participants' tendency to maintain an entrained disfavoured term introduced by an out-community partner compared to an in-community partner.

Table 6. Maintenance \sim 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

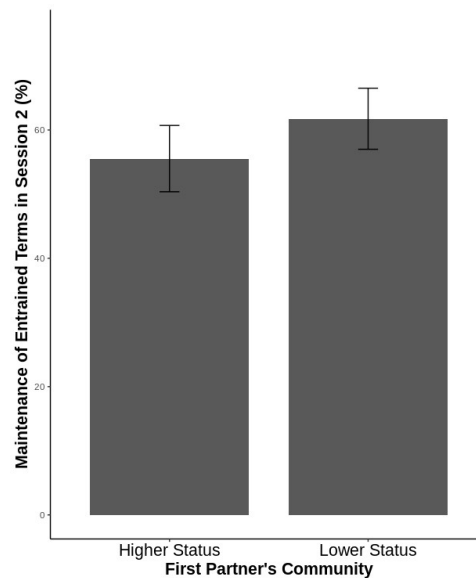


Figure 7. Mean and standard error of percentage of maintenance of disfavoured terms in Session 2 (y-axis) across First Partner's Community (x-axis).

Overall, Experiment 3 found no evidence that speakers' beliefs about the linguistic status of a partner's speech community 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.

General 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 community. We investigated the interplay between individual- 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.

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 disfavoured term for a picture (e.g., *papa* [potato]) influenced participants' subsequent use of that disfavoured term. In Session 2, we measured whether participants would maintain the entrained disfavoured 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., from their own speech community, or Latin American, i.e., from a different community. In Session 1, participants showed significant lexical entrainment, but this was not critically affected by speech community: Use of disfavoured 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 significantly affected by the new partner's speech community. In particular, when participants had entrained to an out-community partner (i.e., a partner from another community) in Session 1, they maintained those terms in

Session 2 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 a new in-community partner than with a new out-community partner.

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 a partner's community in Session 1, but maintenance of entrained terms in Session 2 was critically affected by the new partner's speech community. Importantly, Experiment 3 showed that the differential effects of speaker community on maintenance could not be explained in terms of linguistic status (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) than their own, 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 robust lexical entrainment occurs even when not interacting face-to-face with a partner (e.g., Branigan et al., 2011; Brennan, 1991, 1996; Tobar-Henríquez, Rabagliati & Branigan, 2019). Furthermore, they provide novel evidence that speakers' tendency to maintain previously used referential expressions when interacting with new partners occurs in computer-mediated interactions.

But most importantly, our results provide new insights into the interplay between individual- and community-level influences in language use, and in particular 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 individual-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 individual-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.

Implications for lexical entrainment

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 lexical entrainment 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 speech 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. In sum, there is still significant work to be done in understanding whether, how, and when 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.

From individual experiences to community preferences

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 of which they are not members. More specifically, we suggest that participants showed a tendency to extrapolate out-community partners' disfavoured 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 disfavoured term acted as positive evidence that members of that community preferred that term, leading participants to update their knowledge of that community's preferences, and thus increasing their likelihood to use that term (i.e., maintain its precedent) 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 disfavoured term did not meaningfully 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).

Effects of language status on referential expressions

Our data are also relevant to previous proposals that lexical entrainment (and thus potentially 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 the linguistic status of a partner's language variety. 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 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 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 maintained entrained terms from a lower-status partner and a higher-status partner to the same extent when interacting with a subsequent in-community 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: Whereas 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 disfavoured terms, whereas Experiment 2's (Mexican) participants extrapolated disfavoured terms less often to in-community partners than out-community partners.

Why this difference occurred is uncertain, and could simply reflect measurement error across the studies, but an interesting possibility is that it could be indicative of historical differences in the relationship between Spain and Latin America on the one hand, and Mexico and Argentina on the other. In particular, due to the Spanish colonisation of the Americas, Castilians may assume that Latin Americans are familiar with Castilian lexical preferences (i.e., Castilian favoured terms), thus leading Castilians to maintain disfavoured 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). Thus, Mexicans may have maintained disfavoured Mexican terms more often with Argentinians versus Mexicans as a way to enhance communication with Argentinians. In particular, they might have generalised disfavoured terms across two Argentinians because they were aware of dialectal differences between Mexican and Argentinian varieties of Spanish. 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).

Conclusion

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 beliefs about their 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 individual-level contexts to broader community-level contexts.

References

- 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.
- 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.
- 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.
- Brennan, S. E. (1991). Conversation with and through computers. *User modeling and user-adapted interaction*, 1(1), 67-86.
- Brennan, S. E. (1996). Lexical entrainment in spontaneous dialog. *Proceedings of ISSD*, 96, 41-44.
- 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.
- 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., Yoon, S. O., & Ryskin, R. A. (2015). People as contexts in conversation. *Psychology of learning and motivation*, 62, 59-99.
- 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 His-*

toriografía e Historia de la Lengua Española. Salamanca: Ediciones Universidad de Salamanca, 37-44.

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. Available at <http://dx.doi.org/10.15845/bells.v5i0.688>

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

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1-33). Hillsdale, NJ: Erlbaum.

Clark, E. V. (2007). Young children's uptake of new words in conversation. *Language in Society*, 36(2), 157-182.

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

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

Gregory Jr, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of personality and social psychology*, 70(6), 1231.

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.
- 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.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Lev-Ari, S. (2018). The influence of social network size on speech perception. *Quarterly Journal of Experimental Psychology*, 71(10), 2249-2260.
- Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior research methods*, 43(3), 679-690.
- 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.
- Morett, Sonia (2014). Actitudes lingüísticas en México. Entre el chovinismo y el malinchismo. In 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>.
- 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.
- Palomares, N. A., Giles, H., Soliz, J., & Gallois, C. (2016). Intergroup accommodation, social categories, and identities. In Giles, H. (Ed.), *Communication accommodation theory: Negotiating personal relationships and social identities across contexts*. New York: Cambridge University Press, 123-151.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(2), 169-190.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, 25, 111-163.
- 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*. New York: Cambridge University Press. 205-255.
- 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.
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, 14(5), 779-804.

i 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., $(1 + slope \mid items)$.