Lexical entrainment without conceptual pacts? Revisiting the matching task

Adrian Bangerter, Eric Mayor, Dominique Knutsen

Lexical entrainment without conceptual pacts? Revisiting the matching task

Adrian **Bangerter** (adrian.bangerter@unine.ch), Eric **Mayor** (eric.mayor@unine.ch)
Institute of Work and Organizational Psychology, University of Neuchâtel
Emile Argand 11, 2000 Neuchâtel, Switzerland

Dominique **Knutsen** (dominique.knutsen@univ-lille.fr)
Univ. Lille, CNRS, CHU Lille, UMR 9193 - SCALab - Sciences Cognitives et Sciences Affectives
F-59000 Lille, France

Corresponding author:
Eric Mayor
Institute of Work and Organizational Psychology, University of Neuchâtel
Emile Argand 11,
2000 Neuchâtel, Switzerland
E-mail : eric.mayor@unine.ch

Abstract

Conversational partners who repeatedly refer to the same objects require less and less collaborative effort to do so. This is due to lexical entrainment, the fact that they come to re-use the same words. Lexical entrainment may reflect the elaboration of conceptual pacts, partner-specific agreements about how to name objects which belong to the conversational partners' common ground. Can lexical entrainment occur even if conversational partners cannot develop conceptual pacts about specific objects? In three experiments, we investigated whether lexical entrainment occurs in the matching task even when cards change over trials and partners are not able to develop pacts. We compared two conditions: a classic condition where cards remained the same for each trial, and a new cards condition where cards changed on each trial. Lexical diversity decreased for pairs in the new cards condition (albeit less than for classical pairs); inconsistent reductions in collaborative effort were also observed. Pairs in the new cards condition also were better able to adapt to novel referring situations (involving novel stimuli or new interaction partners) than classic pairs. The results suggest that lexical entrainment in the matching task may be due in part to factors other than the elaboration of conceptual pacts. These may include the development of an overarching meta-perspective on shared features of cards, reflecting category learning processes resulting from reference negotiation.

Keywords: Matching task; collaborative referring; lexical entrainment; conceptual pacts, repeated referring

Declaration of interests: None

Lexical entrainment without conceptual pacts? Revisiting the matching task

Introduction

Conversation is a process whereby people collaborate to make themselves understood to each other (Grice, 1975; H. Clark, 1996). Conversation is coordinated moment-by-moment on multiple levels. At a very basic level, people try to take turns speaking in the most efficient way possible (Holler, Kendrick, Casillas, & Levinson, 2016; Sacks, Schegloff, & Jefferson, 1974). Another aspect of conversation that gets coordinated is meaning. For example, conversational partners might need to coordinate a specific way of referring to a house where they are to meet the next day. Conversational partners who repeatedly refer to the same objects require less and less collaborative effort (i.e., fewer words, fewer turns) to do so (H. Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987). This is due to lexical entrainment, the fact that they come to re-use the same words when referring to the same objects again (Brennan & H. Clark, 1996). Lexical entrainment may reflect the elaboration of conceptual pacts, which are partner-specific agreements about how to name an object that is part of the common ground of conversational partners.

Can lexical entrainment occur even if conversational partners cannot develop conceptual pacts? By repeatedly talking together, conversational partners learn about each other’s ways of speaking or interactional habits (Schober & Carstensen, 2010). Moreover, the precedents elaborated as a natural outcome of conversational interaction may constitute resources for facilitating reference to novel referents. As a result, partners with a shared conversational history might find it easier to refer to novel objects for which they have no conceptual pacts than partners with no such history. The evidence for this possibility is mixed. A study where married couples and couples composed of strangers of different ages referred to familiar and unfamiliar referents found that married couples were no more efficient in referring than strangers (Schober & Carstensen, 2010). Similar studies (Fussell & Krauss, 1989; Boyle, Anderson, & Newlands, 1994; Gann & Barr, 2014) also found little evidence for a benefit of shared conversational history in referring to novel objects. On the other hand, the conceptual pacts established in referential communication can lead to conceptual changes in participants that facilitate the reuse of conceptual pacts to describe novel objects. In one study, pairs collaborating to build Lego models of cars developed partner-specific conceptual pacts to describe the various pieces they used. When subsequently building a similar model, they extended the existing labels to describe novel pieces (Markman & Makin, 1998). In another study, establishing conceptual pacts broadened participants' attention to functional features of stimuli and enhanced category learning (Voiklis & Corter, 2012). Finally, there is also evidence for lexical differentiation (Van Der Wege, 2009) as a complementary process to lexical entrainment, whereby participants expand or modify previously established pacts to accommodate novel objects of reference that are similar to previously discussed objects. Taken together, it seems plausible that participants with a shared conversational history may be able to use that history to facilitate referring to novel objects. However, it is unclear what role, if any, the establishment of conceptual pacts plays in such a putative process.

Much of the evidence for the establishment of conceptual pacts has been derived from the matching task paradigm. In the classic version of this paradigm (H. Clark & Wilkes-Gibbs, 1986), two participants are each given a set of identical cards with ambiguous pictures (there is no straightforward and shared label to describe the picture; tangram figures are often used). One participant, the director, has the set in a correct order, and the other, the matcher, has the set in a random order. The task consists in rearranging the matcher's set such that the cards are in the same order as the director's. To accomplish this, participants need to identify each card and thus to describe the cards to each other (participants cannot see each other, nor can they see their partner's cards). Once the task has been completed, the cards are shuffled, a new correct order is specified for the director, and participants complete the task again (typically over 4 to 6 iterations with the same participants and cards). Participants require a lot of collaborative effort in the initial trials of the matching task. This is because they need to agree on a conceptual perspective to describe each card, and to do this they need to reach the mutual understanding that they are talking about the same card. To achieve this, they suggest tentative characterizations of the pictures and revise or expand those characterizations in interactive sequences to which the matcher often contributes significantly (Tolins, Zeamer, & Fox Tree, 2017). Once they have established a conceptual pact about how to refer to that card, in subsequent trials, they often reuse the label corresponding to the pact. Subsequent referring thus simply consists of the director mentioning the label of the corresponding card and the matcher acknowledging that the card has been identified and placed via a back-channel utterance (Bangerter & H. Clark, 2003). As a result, they require much less collaborative effort to complete the task (H. Clark & Wilkes-Gibbs, 1986). To illustrate this process, here is an example of six successive references to the same card (a tangram figure) by a director (H. Clark & Wilkes-Gibbs, 1986, p. 12):

1. All right, the next one looks like a person who's ice skating, except they're sticking two arms out in front.
2. Um, the next one's the person ice skating that has two arms?
3. The fourth one is the person ice skating, with two arms.
4. The next one's the ice skater.
5. The fourth one's the ice skater.
6. The ice skater.

The matching task has a long heritage (for an early version, see Krauss & Weinheimer, 1966) and can be described as the workhorse experimental task for studying collaborative referring in conversation. Many variations exist. It has been used to study how experts and novices adapt to each other (Isaacs & H. Clark, 1987) and with native and non-native speakers (Bortfeld & Brennan, 1997). Other variations have examined the role of the matcher. One study used a third matcher unable to contribute to the grounding process to study the effects of collaborative referring on overhearers (Schober & H. Clark, 1989). Another used new matchers more or less aware of the conversation between the director and first matcher (Wilkes-Gibbs & H. Clark, 1992). Yet another tested a director doing the task simultaneously with two independent matchers to study the limits of conceptual pacts (Horton & Gerrig, 2002). The matching task is amenable to manipulation of the stimulus pictures (e.g., their codability may vary; Hupet, Seron, & Chantraine, 1991) and can be used to study multimodal referring, i.e., the use of gestures in constructing referring expressions in mutually visible pairs (De Ruiter, Bangerter, & Dings, 2012). Even more recently, variations of the matching task have been used to compare situations in which dialogue partners discuss one versus several perspectives before reaching an agreement as to which reference should be used (Knutsen, Ros, & Le Bigot, 2018).

Repeatedly replicated findings from the matching task have been taken as evidence of the general principle of audience design (H. Clark & Murphy, 1982), namely that speakers take into account their common ground – that is, the knowledge that they share and are aware of sharing (H. Clark, 1996) – to design utterances that their current addressees are capable of understanding easily. Researchers agree that lexical entrainment via conceptual pacts is the primary factor explaining the progressive decrease in collaborative effort necessary for pairs of participants to complete the matching task over several trials (Brennan & H. Clark, 1996; H. Clark & Wilkes-Gibbs, 1986). In recent years, however, the question of what the matching task, at least in its classic, relatively unconstrained version, shows and does not show has become somewhat controversial. One issue was raised by Keysar (1997), who suggested that many experiments purporting to show the role of common ground in language use actually fail the criterion of parsimony, because they are unable to distinguish between cases where something that is mutually known to the speaker and addressee is used to design utterances *because* it is mutually known and cases where it is used because it is simply more salient or accessible to the speaker. The classic matching task is a case in point, because while lexically entrained expressions like *the ice skater* may get used because they are mutually known, they may also get used simply because they are easily accessible to the speaker (indeed, participants are more likely to reuse expressions they have produced themselves than those produced by their partners; Knutsen & Le Bigot, 2014).

Another issue concerns the extent to which coordination in dialogue derives purely or even mainly from coordination of referring expressions, i.e., from lexical entrainment. Mills (2014, p. 159) suggested that “experimental approaches that do study the emergence of conventions in dialogue typically restrict their analyses to the study of referring conventions”, and tend to neglect the potential contribution of interactional routines to coordinated, efficient dialogue. Mills distinguished between *semantic* and *procedural* coordination. Semantic coordination concerns the well-documented effects of lexical entrainment via conceptual pacts. Procedural coordination concerns the (much less studied) effects of coordination of how to make one’s contribution, how to progress from one step of the task to the next (Bangerter & H. Clark, 2003), and the like. For example, using the classic maze game paradigm, Garrod and Anderson (1987) found that participants converged on different interactive routines for describing their positions in a maze to each other. Fusaroli and Tylén (2016) found that gradually emerging structural organization during interactions improved conversational performance. In a recent study using conversational data from Experiment 1 of this study, Knutsen, Bangerter, and Mayor (2019) found that procedural speech in the matching task could be divided into two categories: specific procedural coordination (which includes efforts specific to the matching task, e.g., negotiating the director/matcher roles, deciding in which order the cards should be described, etc.) and generic procedural coordination (which includes efforts which must be made in all joint tasks, such as using markers to coordinate speech turn-taking). These two aspects of procedural coordination constituted almost 30% of the total amount of speech in the matching task. Importantly, the results of this study suggest that just like conceptual pacts, procedural agreements also contribute to reducing the partners' collaborative efforts (see also Bangerter & Clark, 2003). Indeed, once conversational partners have reached an agreement as to how to perform the task, they do not need to renegotiate this aspect of the task again in later trials, resulting in a decrease in the number of words and speech turns necessary to reach mutual understanding. For instance, the participants may initially describe the spatial orientation of how the cards are placed (*a grid with two rows of four cards each*). In subsequent rounds, they may refer implicitly to previously grounded descriptions when coordinating on the next step in the task (*the next one is the first card in the second row*).Therefore, it seems that gains in collaborative efficiency are not due to purely semantic coordination, but also to procedural coordination.

Aside from these issues, even if we focus exclusively on semantic coordination *sensu* Mills (2014, it still seems unclear to what extent collaborative gains over time in the matching task reflect the sole influence of conceptual pacts and their effects on lexical entrainment. Indeed, collaborative gains may reflect conceptual pacts, but they may also emerge from participants' shared conversational history. In the classic version of the task, both potential effects are confounded. In the above example, when the director initially says *the next one looks like a person who's ice skating*, the conceptualization of the tangram figure as an *ice skater* may develop into a conceptual pact as described by previous research (Brennan & H. Clark, 1996). But it may also facilitate reference to similar referents encountered for the first time in the same or later trials. For instance, it may encourage directors to describe the pictures as human figures instead of animals or even geometrical shapes. Thus, efficient performance in the matching task may potentially arise from other cognitive and coordinative processes the potential benefits of which may emerge over time. This makes sense given the multilayered nature of conversational interaction (H. Clark, 1996). In interacting together, participants coordinate on a range of aspects; precedents established on any of these aspects may potentially contribute to better collaborative performance.

In its classic form, then, the matching task confounds the effects of a shared interactional history with those of the elaboration of conceptual pacts, because participants (typically zero-acquaintance, Schober & H. Clark, 1989) repeatedly refer to the same objects while their interactional history develops. Moreover, because they refer to the same objects, there is no opportunity to test the potential benefits of referring to novel objects. We therefore modified the matching task in order to investigate these issues. We contrasted the classic matching task with a novel condition where participants complete the task with new cards on each trial. Participants who deal with new cards on each trial accumulate a shared interactional history in much the same way as participants who accomplish the classic matching task, but are not able to establish and use conceptual pacts. If the standard phenomena observed in the matching task and typically attributed to the effects of the establishment of conceptual pacts are also observable to some degree in a version of the task where participants are confronted with new cards on each trial, then it is possible to estimate the relative impact of conceptual pacts versus a shared interactional history on the facilitation of reference to novel objects.

Revisiting the Matching Task: The Current Experiments

In the current experiments, pairs of participants completed the matching task either in the *classic* condition (where the cards remain the same on each trial) or a version of the task where participants place new cards on each trial (hereafter the *new cards* condition). In the classic condition, repeated reference to the same objects is expected to lead to the establishment of conceptual pacts. In the literature on collaborative reference (Brennan & H. Clark, 1996; H. Clark & Wilkes-Gibbs, 1986), this is observable via three main outcome variables. First, over trials, *collaborative effort* to complete the task decreases. This is typically measured by the number of words or turns produced by both participants in total per trial. In our experiment, we measured it by the number of words used by both participants to describe cards. Second, over trials, lexical entrainment occurs (this is part of the cause for the reduction in collaborative effort), meaning that participants come to use similar terms to describe the cards. As a result, *lexical diversity* (in our experiments, the proportion of new lexical items used to describe cards relative to the total number of different lexical items used to describe ecards) decreases. Third, the establishment of a conceptual pact reflects a partner-specific precedent about how to refer to an object. It thus entails mutual expectations that it will be used in the future to describe that object (Brennan & H. Clark, 1996). One way participants mark this is via the use of definite reference to indicate that the object being described is mutually known (*the ice skater* replaces *an ice skater*). We measured this outcome by computing the opposite phenomenon, a decrease in *indefinite reference*. Decrease in indefinite reference (e.g., *a person*) is less ambiguous to code automatically than an increase in definite reference (*the person*) because definite pronouns may be used in more contexts than those of referring expressions.

We expected pairs in both conditions to exhibit a decrease in collaborative effort (measured by a decrease in words) and a corresponding increase of lexical entrainment (measured by a decrease in lexical diversity). However, we expected pairs in the new cards condition to show both phenomena to a lesser degree than pairs in the classic condition. We also expected pairs in the new cards condition to rely less on conceptual pacts in repeated referring than pairs in the classic condition (measured by a higher rate of indefinite reference).

For all of these analyses, it is important to point out that, like any joint activity, matching task conversations also feature procedural coordination (Mills, 2014). That is, because up to 30% of words produced in the matching task may serve procedural coordination instead of describing card features (Knutsen et al., 2019), it is important to take this into account. To get a better estimate of the effects of conceptual pacts versus a shared interactional history on the three main variables above, we removed all words relative to card placement coordination. This consists of utterances coordinating which card to identify and place, e.g., “third slot in the top row” or “let’s start with”. In other words, the words analyzed in the current experiments have been “cleaned” of procedural talk about how to place the cards, because we are interested in a clean comparison of the effects of conceptual pacts versus a shared interactional history on card descriptions.

We conducted three experiments with variations on this basic design. In Experiment 1, we simply compared the new cards condition with the classic condition. Pairs completed 5 trials with eight-card sets. In Experiment 2, we replicated Experiment 1 and added a sixth trial where pairs in both conditions received new cards. Thus, pairs in the classic condition were faced with novel cards for the first time, whereas pairs in the new cards condition simply continued as in Trials 1-5. This design tests whether pairs exposed to a wider variety of objects and repeatedly forced to adapt their communication are more flexible than pairs that have repeatedly coordinated via a limited set of conceptual pacts. This conjecture follows from the fact that negotiating reference has conceptual implications changing the way participants attend to and think about stimuli (Voiklis & Corter, 2012). We expected classic pairs to perform less well on Trial 6 than new-cards pairs. In Experiment 3, pairs in both conditions initially performed 4 trials. We then switched the matcher in each pair for a second block of 4 trials. This design allowed testing whether collaborative gains in each condition can be transferred from a pair member to a new pair. Whenever necessary, specific aspects of the rationale of each experiment are described in the introductory section of that experiment. The three experiments were carried out in accordance with the Declaration of Helsinki. All participants signed an informed consent form. All the stimuli, data and scripts used in each experiment are available at https://osf.io/a4m7k/

Experiment 1

Method

Participants. We recruited participants (*N* = 44 native French speakers, 25 women) from the student body of a Swiss university. They received 10 CHF for participating. Unacquainted participants were scheduled to arrive in pairs so that they could complete the task together. They were randomly allocated to either the director or matcher role. Pairs were randomly allocated to either the classic condition or the new cards condition. As we expected more variance in the new cards condition, 14 pairs were allocated to it, whereas 8 dyads were allocated to the classic condition.

Materials and procedure. Members of a pair sat next to each other in front of a different computer screen. A separation was placed so that participants could not see each other's screen nor their partner during the experiment. Participants arranged a set of eight cards depicting humanoid tangram shapes. The shapes were chosen to be similar to those used in previous matching task experiments. These were displayed to participants on a computer screen in two rows of four columns using a program we developed in Flash (Action Script). After a practice trial with eight cards featuring everyday objects, on each of 5 trials, both participants viewed the set of eight tangram cards on their screen (the same set for both of them but in a different order). While the directors' view included cards already in placement slots (a grid with two rows of four cards), matchers saw their placement slots displayed above the cards. They moved cards to the slots by clicking and dragging them. They could move cards back and forth freely. Directors' cards could not be moved. The experiment used a pool of 40 different cards. In the classic condition, participants arranged the same set of cards on each trial (cards were drawn randomly from the pool and the order of the cards was randomized at each trial). In the new cards condition, eight cards were drawn without replacement from the pool on each trial. Pairs completed the entire experiment without feedback as to the accuracy of their performance and were subsequently debriefed.

Data preparation. We audio-recorded participants' talk and transcribed the recordings verbatim. Then, all talk relative to the placement of cards was removed (see Knutsen et al., 2019, for a description of the coding procedure and reliability). Transcripts were then coded for each dependent variable (collaborative effort, lexical diversity and indefinite reference) per card and per trial. Word count was our measure of collaborative effort and was computed per card per trial (both participants together). Indefinite reference was also computed automatically (by counting the frequency of the French pronouns *un*, *une* and *des* and dividing by the word count) per card per trial. Lexical diversity was operationalized as a variation on typical type-token ratio measures (Richardson, 1987). We computed the proportion of new word types (nouns or verbs, as these are the most frequently used word classes in the matching task) used to describe a given card on a given trial, i.e., the number of word types that were not used to describe any cards on previous trials, divided by the total number of word types used to describe the card on that trial. Therefore, lexical diversity always had a value of 1 at the first trial (all word types used are new), and a value ≤ 1 on subsequent trials. We manually identified verbs and nouns used to describe each card per trial (interrater agreement, computed on a subset of the data (20 trials) double-coded by two coders, was acceptable, $r = .75$). The data was then passed to an R script which automatically computed the ratio of new word types to total word types.

Data analysis. For all dependent variables, data were analyzed using linear mixed-effects models (Baayen, Davidson & Bates, 2008; Barr et al., 2013). These analyses were conducted using SAS 9.4 (GLIMMIX procedure). We included Trial (linear and quadratic trends) and Condition as fixed effects. Following Barr et al. (2013), we started by running analyses which included the maximal random effects structure (i.e., they included all random slopes and random intercepts justified by the experimental design; in the current study, this would include by-dyad and by-item random intercepts, by-item random slopes corresponding to Condition and by-dyad and by-item random slopes corresponding to Trial (linear and quadratic trends)). However, models including the maximal random effects structure often fail to converge. When this happened, we identified the random effects causing convergence issues (this is performed automatically in SAS; see McMahon, Pouget, & Tortu, 2006). We then removed these effects and ran the analysis again (this does not affect the outcome of the models; it only affects how the degrees of freedom are calculated). The results reported hereafter are the results of the second analysis; the random effects which were finally included in each analysis are listed for each model.

For all dependent variables, we first tested a model (hereafter Model 1) including Trial (linear and quadratic trends), Condition and their interaction terms as predictors. This model allows examining whether Trial and Condition affect the dependent variables and whether the effects of Trial differ significantly by Condition. For these analyses, estimates for each predictor represent its unique effect (i.e., controlling for the effects of other predictors).

In order to make significant interactions easier to interpret, we then separately considered the data from the classic condition (in Model 2) and the new cards condition (in Model 3). Both Model 2 and Model 3 included the linear and quadratic trends of Trial in each condition.

In analyses of lexical diversity, we excluded Trial 1 from the analyses in Models 1-3 because its value is always 1 (a constant). Thus, those analyses were ran on Trials 2-5 for lexical diversity. Given that lexical diversity is likely to decrease strongly between Trials 1 and 2, an additional analysis was conducted in which we tested if lexical diversity at Trial 2 was significantly different from 1 (separately for both conditions), by subtracting 1 from lexical diversity, and predicting the resulting variable in a mixed model which included no fixed intercept. In such a model, the coefficients for each factor level correspond to the difference between this level and a value of 0. However, because the value of 1 was initially subtracted, this model actually tests the difference between the level and a value of 1. A significant coefficient would therefore imply that the level of the factor significantly differs from 1.

Results

Figure 1 presents the means by Trial and Condition (and standard errors of the means) for all three dependent variables.

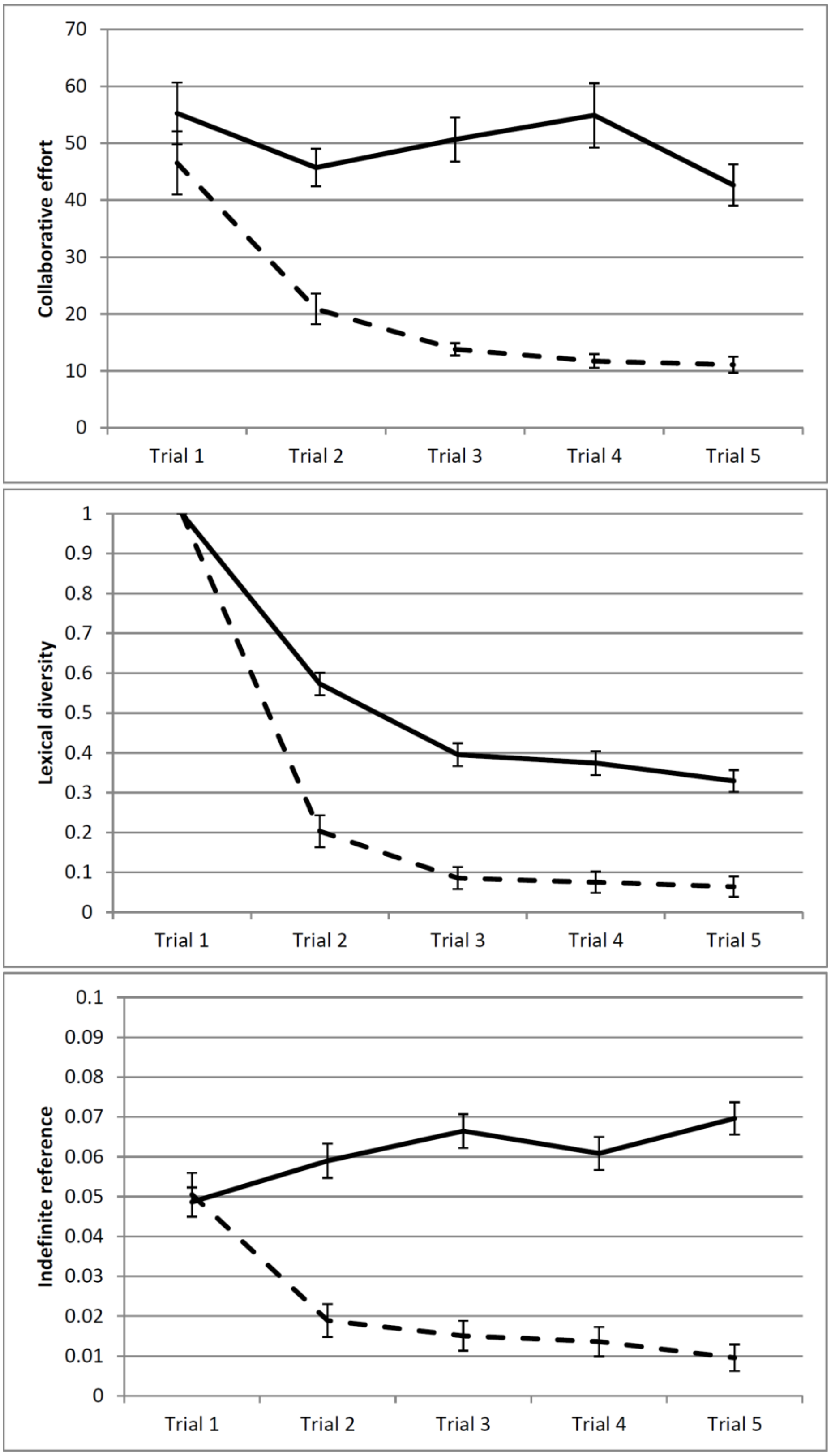

Figure 1: Collaborative effort (number of words per figure), lexical diversity (proportion of new lexical items divided by the total number of words used) and indefinite reference (number of indefinite references divided by the total number of words used) by trial and condition for Experiment 1. Dashed lines represent the classic condition. Plain lines represent the new cards condition. Error bars represent the standard error of the mean.

Collaborative effort. Table 1 displays the results for Models 1, 2 and 3. The significant negative linear trend for Trial in Model 1 shows that collaborative effort decreased in the classic condition. The significant positive quadratic trend shows that this decrease was less important in later trials in the classic condition. The effect of Condition was non-significant, suggesting that the amount of collaborative effort was similar in both conditions (controlling for Trial). A significant linear trend–by-Condition interaction and a significant quadratic trend-by-Condition interaction were also found. The effects of Trial were investigated further in Models 2 and 3. These revealed a significant negative linear trend and a significant positive quadratic trend in the classic condition, confirming the results from Model 1. They also revealed that neither trend was significant in the new cards condition. Thus, collaborative effort did not decrease over trials in the latter condition.

Table 1

Experiment 1 – Results of the Analysis on Collaborative Effort

Model	**Model structure and coefficients**
Model 1 (All data)	
Random effects included	By-dyad and by-item intercepts; by-item slopes corresponding to Condition
Effect of linear trend	$F(1, 818) = 10.56$, $p = .001$, $b = -31.56$, $SE = 7.26$
Effect of quadratic trend	$F(1, 818) = 5.19$, $p = .023$, $b = 3.93$, $SE = 1.19$
Effect of Condition	$F(1, 194) = 2.32$, $p = .129$, $b = -21.66$, $SE = 14.22$
Linear trend-by-Condition interaction	$F(1, 818) = 13.09$, $p < .001$, $b = 33.26$, $SE = 9.19$
Quadratic trend-by-Condition interaction	$F(1, 818) = 8.69$, $p = .003$, $b = -4.43$, $SE = 1.50$
Model 2 (classic condition only)	
Random effects included	By-dyad and by-item intercepts; by-dyad slopes corresponding to the linear trend
Effect of linear trend	$F(1, 279) = 52.11$, $p < .001$, $b = -31.56$, $SE = 4.37$
Effect of quadratic trend	$F(1, 235) = 31.03$, $p < .001$, $b = 3.93$, $SE = 0.70$
Model 3 (new cards condition only)	
Random effects included	By-dyad and by-item intercepts
Effect of linear trend	$F(1, 535) = 0.05$, $p = .820$, $b = 1.50$, $SE = 6.60$
Effect of quadratic trend	$F(1, 535) = 0.17$, $p = .676$, $b = -0.45$, $SE = 1.08$

Lexical diversity. Lexical diversity at Trial 2 was lower than the theoretical value of 1 in both conditions (in the classic condition: $b = -0.79$, $SE = 0.03$, $p < .001$; in the new cards condition: $b = -0.43$, $SE = 0.04$, $p < .001$). Results for Trials 2-5 are provided in Table 2. The significant negative linear trend for Trial in Model 1 shows that lexical diversity decreased in the classic condition. The significant positive quadratic trend shows that this decrease was less important in later trials in the classic condition. The significant effect of Condition shows that lexical diversity was lower in the classic condition than in the new cards condition (controlling for Trial). Neither the linear trend-by-Condition interaction nor the quadratic trend-by-Condition interaction were significant. The effects of Trial were investigated further in Models 2 and 3. These revealed non-significant linear and quadratic trends in the classic condition, contrary to what was found in Model 1. Thus, lexical diversity did not decrease over Trials 2-5 in the classic condition. They also revealed a significant negative linear trend and a significant quadratic trend in the new cards condition, suggesting that lexical diversity decreased over Trials 2-5 in that condition.

Table 2

Experiment 1 – Results of the Analysis on Lexical Diversity

Model	**Model structure and coefficients**
Model 1 (All data)	
Random effects included	By-dyad and by-item intercepts; by-item slopes corresponding to Condition and the quadratic trends
Effect of linear trend	$F(1, 626) = 11.06$, $p = .001$, $b = -0.20$, $SE = 0.12$
Effect of quadratic trend	$F(1, 634) = 6.63$, $p = .010$, $b = 0.02$, $SE = 0.02$
Effect of Condition	$F(1, 646) = 4.35$, $p = .037$, $b = 0.51$, $SE = 0.25$
Linear trend-by-Condition interaction	$F(1, 627) = 0.40$, $p = .526$, $b = -0.10$, $SE = 0.15$
Quadratic trend-by-Condition interaction	$F(1, 632) = 0.16$, $p = .686$, $b = 0.01$, $SE = 0.02$
Model 2 (classic condition only)	
Random effects included	By-dyad and by-item intercepts
Effect of linear trend	$F(1, 216) = 3.52$, $p = .062$, $b = -0.20$, $SE = 0.11$
Effect of quadratic trend	$F(1, 216) = 2.27$, $p = .133$, $b = 0.02$, $SE = 0.02$
Model 3 (new cards condition only)	
Random effects included	By-dyad and by-item intercept; by-dyad and by-item slopes corresponding to linear and quadratic trends
Effect of linear trend	$F(1, 423) = 9.78$, $p = .002$, $b = -0.30$, $SE = 0.10$
Effect of quadratic trend	$F(1, 425) = 5.56$, $p = .019$, $b = 0.03$, $SE = 0.01$

Indefinite reference. Table 3 displays the results for Models 1, 2 and 3. The significant negative linear trend for Trial in Model 1 shows that the production of indefinite references decreased in the classic condition. The quadratic trend was non-significant, suggesting that the decrease remained constant across trials. The significant effect of Condition shows that fewer indefinite references were produced in the new cards condition than in the classic cards condition (controlling for Trial). A significant linear trend-by-condition interaction and a significant quadratic trend-by-condition interaction were also found. The effects of Trial were investigated further in Models 2 and 3. These revealed a significant negative linear trend and a significant positive quadratic trend in the classic condition, confirming and extending the results from Model 1. They also revealed that only the linear trend was significant in the new cards condition. Importantly, the linear trend was positive rather than negative, showing that indefinite reference production actually increased across trials in this condition (as shown in the lower panel of Figure 1).

Table 3
Experiment 1 – Results of the Analysis on Indefinite Reference

Model	Model structure and coefficients
Model 1 (All data)	
Random effects included	By-dyad and by-item intercepts; by-item slopes corresponding to Condition and the linear and quadratic trends
Effect of linear trend	$F(1, 692) = 4.59$, $p = .033$, $b = -0.03$, $SE = 0.01$
Effect of quadratic trend	$F(1, 754) = 2.93$, $p = .088$, $b < 0.01$, $SE < 0.01$
Effect of Condition	$F(1, 308) = 7.77$, $p = .006$, $b = -0.03$, $SE = 0.01$
Linear trend-by-Condition interaction	$F(1, 806) = 24.67$, $p < .001$, $b = 0.05$, $SE = 0.01$
Quadratic trend-by-Condition interaction	$F(1, 806) = 13.51$, $p < .001$, $b = -0.01$, $SE < 0.01$
Model 2 (classic condition only)	
Random effects included	By-dyad and by-item intercepts
Effect of linear trend	$F(1, 286) = 27.47$, $p < .001$, $b = -0.03$, $SE = 0.01$
Effect of quadratic trend	$F(1, 286) = 15.56$, $p < .001$, $b < 0.01$, $SE < 0.01$
Model 3 (new cards condition only)	
Random effects included	By-dyad and by-item intercepts; by-item slopes corresponding to the linear trend
Effect of linear trend	$F(1, 540) = 5.73$, $p = .017$, $b = 0.01$, $SE = 0.01$
Effect of quadratic trend	$F(1, 537) = 3.10$, $p = .079$, $b < -0.01$, $SE < 0.01$

Discussion

Pairs in the classic condition showed the signature pattern of lexical entrainment through conceptual pacts: A reduction of lexical diversity, decreasing use of indefinite reference suggesting the use of conceptual pacts, and a corresponding reduction in the collaborative effort necessary to complete the task. Pairs in the new cards condition showed a reduction in lexical diversity, but also an unexpected *increase* in indefinite reference (the increase is unexpected because there is no corresponding increase in novelty over trials), but did not decrease collaborative effort.

That new cards pairs reduced lexical diversity is a noteworthy finding, suggesting that repeated referring to the same objects is not a necessary condition for lexical entrainment. Participants were unable to develop and use conceptual pacts to refer to the same objects. However, given that the figures they placed each trial came from a similar universe of objects (i.e., anthropoid tangrams) as those on the previous trial, they may have implicitly entrained on a more overarching meta-perspective, like the assumption that the figures are all humanoid. Alternatively, they may have been able to reuse and differentiate (Van der Wege, 2009) terms used to refer to previous figures when encountering a similar figure in a later trial. Whatever they did, the results clearly show that lexical entrainment can take place outside of the possibility to establish conceptual pacts by repeated referring to the same objects. An example of this could be the fact that a new cards pair describes a diamond shape as the “head” of a figure on Trial 1. Confronted with a similar anthropoid card on Trial 2, that same pair could simply refer to a head without having to entrain on the perspective that the figures have heads. The lack of a significant collaborative effort decrease for new cards pairs is thus somewhat surprising, given that they reduced lexical diversity and given previous findings (Markman & Makin, 2008) suggesting that partners are able to extend previously used labels to new, similar objects. This may have been due to the larger variability in the new cards condition. While some pairs were able to decrease collaborative effort systematically, others had difficulties doing so. Indeed, while all 8 classic pairs decreased collaborative effort between Trial 1 and Trial 5, only 9 out of 14 new cards pairs were able to do so. While classic pairs progressively reduced the uncertainty about card identification, new cards pairs remained vulnerable to the risk of a new card unexpectedly posing difficulties to describe and thus requiring more collaborative effort. The unexpected increase in indefinite reference in the new cards condition is puzzling. Given these somewhat unexpected results, in Experiment 2, we designed a replication and extension.

Experiment 2

In Experiment 2, we wanted to test potential effects of the different shared experiences that classic and new cards pairs develop. According to earlier research on category coordination in communication (Markman & Makin, 2008; Voiklis & Corter, 2012), communication coordinates category structures between participants, possibly by developing perspectives that focus their attention on salient or relevant properties of objects (E. Clark, 1997). If so, then participants who communicate about a wide range of objects (new cards pairs) may develop more general and thus more robust category structures or more broad perspectives (E. Clark, 1997) than participants who repeatedly refer to the same objects (classic pairs). This in turn might make them better able to communicate about novel objects. We thus had pairs in both conditions complete a sixth trial immediately after the fifth, where they received a set of new cards. This amounts to more of the same for pairs in the new cards condition, but places pairs in the classic condition in a novel situation. We compared collaborative effort, lexical diversity and indefinite reference on Trial 6 between conditions to test whether either condition performed better than the other or not. We also compared the differences between Trials 5 and 6 for each condition in order to test how performance changed within-condition. Finally, we tested whether collaborative effort and indefinite reference differed between Trial 1 and Trial 6 for each condition, to determine whether performance at Trial 6 was different from initial performance without any common ground.

Moreover, we wanted to test whether (1) we could replicate the decrease in lexical diversity in the new cards condition, (2) whether the finding that collaborative effort does not significantly decrease is found again, and (3) whether indefinite reference would increase again in the new cards condition (a finding we did not expect) or not.

Method

Participants. We recruited participants (*N* = 60 native French speakers) from the student body of a Swiss university. Participants completed the experiment in pairs (15 pairs in each condition; this time we chose to increase the sample size to better be able to deal with potentially large variances in the new cards condition) in an identical fashion as for Experiment 1. They received 10 CHF for participating.

Materials and procedure. The materials were similar to those used in Experiment 1. An additional eight cards were added to the pool of forty cards from Experiment 1. The procedure was the same as for Experiment 1 with the exception of the additional sixth trial. On Trial 6, pairs in both conditions received 8 new cards. For new cards pairs, those cards were whichever eight cards were remaining in the pool. For classic pairs, those cards were a randomly determined set of eight cards different from the cards they had placed for Trials 1-5.

Data preparation. Like in Experiment 1, we audio-recorded participants' talk and transcribed the recordings verbatim. Transcripts were then coded for each dependent variable (collaborative effort, lexical diversity and indefinite reference) per card and per trial following the same procedure as in Experiment 1.

Data analysis. For all dependent variables, data from Trials 1-5 were analyzed as in Experiment 1 (that is, we generated Models 1, 2 and 3 again for each dependent variable; an additional analysis was conducted to examine the decrease in lexical diversity between Trials 1 and 2). Moreover, for lexical diversity, we also tested the difference between Trial 6 and a theoretical value of 1.

We further tested differences between conditions on Trial 6 by generating, for each DV, an additional model that included Trial as a factor variable (rather than as linear and quadratic trends), Condition and their interaction term. In this analysis, we included the main effect of trial with Trial 6 as the reference category; the effect of condition therefore represents the difference between conditions at Trial 6. We also tested the differences between Trials 5 and 6 in both conditions separately, as well as differences between Trials 1 and 6 in both conditions separately (except for lexical diversity).

Results

Figure 2 presents the means by Trial and Condition (and standard errors of the means) for all three dependent variables in Experiment 2.

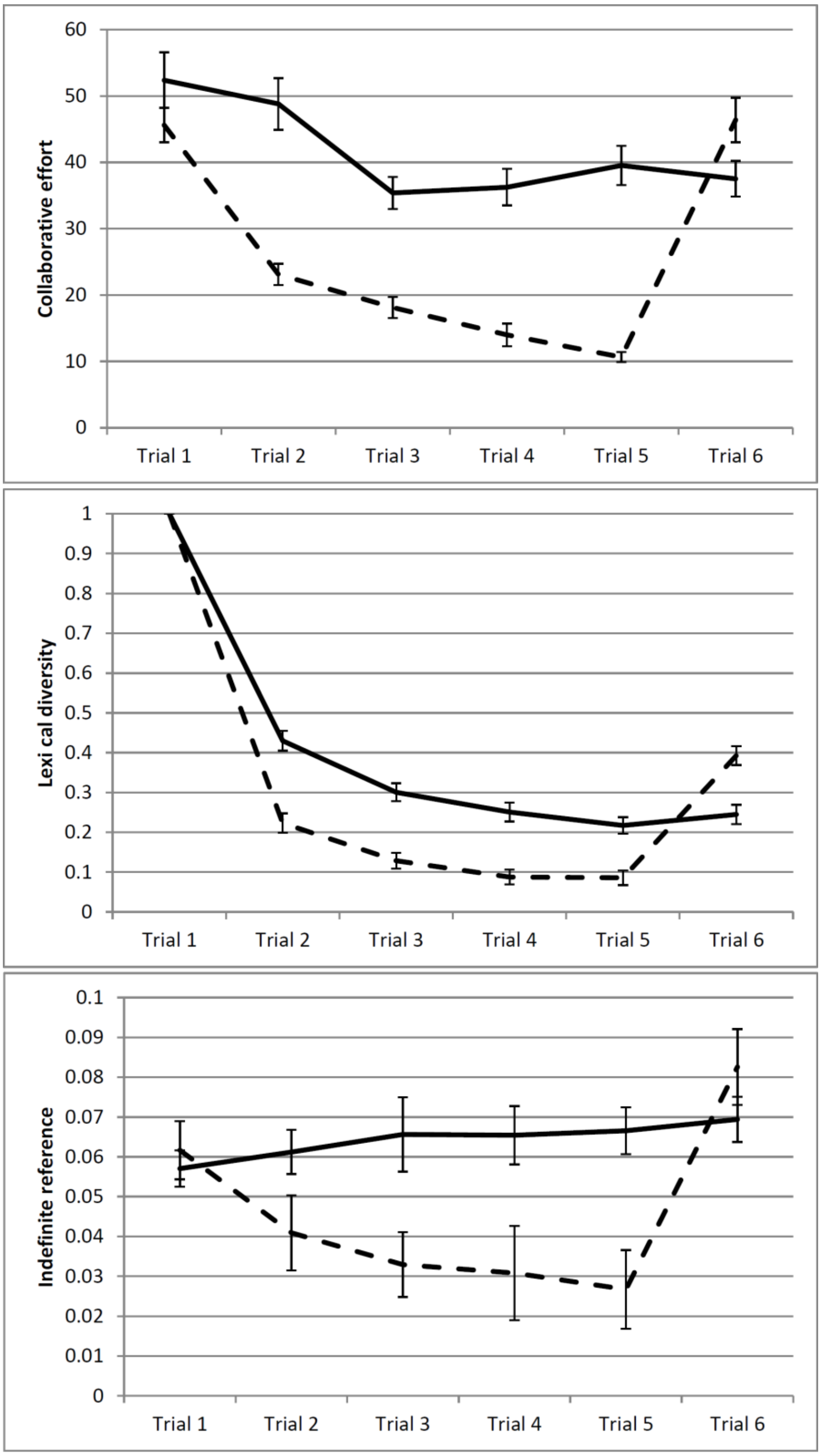

Figure 2: Collaborative effort (number of words per figure), lexical diversity (proportion of new lexical items divided by the total number of words used) and indefinite reference (number of indefinite references divided by the total number of words used) by trial and condition for Experiment 2. Dashed

lines represent the classic condition. Plain lines represent the new cards condition. Error bars represent the standard error of the mean.

Collaborative effort. Table 4 displays the results for Models 1, 2 and 3 (Trials 1 – 5 only). The significant negative linear trend for Trial in Model 1 shows that collaborative effort decreased in the classic condition. The significant positive quadratic trend shows that this decrease was less important in later trials in the classic condition. The effect of Condition was non-significant, suggesting that the amount of collaborative effort was similar in both conditions (controlling for Trial). Neither the linear trend-by-Condition interaction nor the quadratic trend-by-Condition interaction were significant. The effects of Trial were investigated further in Models 2 and 3. These revealed a significant negative linear trend and a significant positive quadratic trend in the classic condition, confirming the results from Model 1. They also revealed a significant negative linear trend and a significant positive trend in the new cards condition, also confirming the results from Model 1.

The additional analysis revealed that collaborative effort at Trial 6 was greater in the classic condition than in the new cards condition, $F(1, 29) = 21.09$, $p < .001$. In the classic condition, collaborative effort at Trial 6 was not significantly different from collaborative effort at Trial 1, $b = -0.47$, $SE = 4.75$, $p = .922$. Moreover, participants required more collaborative effort at Trial 6 than at Trial 5, $b = -35.63$, $SE = 4.84$, $p < .001$. In the new cards condition, participants required less collaborative effort at Trial 6 than at Trial 1, $b = 13.72$, $SE = 4.73$, $p = .005$. Moreover, collaborative effort at Trial 6 was not significantly different from collaborative effort at Trial 5, $b = 2.28$, $SE = 4.69$, $p = .628$.

Table 4

Experiment 2 – Results of the Analysis on Collaborative Effort (Trials 1 – 5)

Model	**Model structure and coefficients**
Model 1 (All data)	
Random effects included	By-dyad and by-item intercepts; by-dyad slopes corresponding to the linear trend; by-item slopes corresponding to Condition
Effect of linear trend	$F(1, 1087) = 49.50$, $p < .001$, $b = -24.89$, $SE = 4.10$
Effect of quadratic trend	$F(1, 1087) = 26.98$, $p < .001$, $b = 2.85$, $SE = 0.67$
Effect of Condition	$F(1, 555) = 0.03$, $p = .854$, $b = 1.47$, $SE = 8.00$
Linear trend-by-Condition interaction	$F(1, 1087) = 2.39$, $p = .122$, $b = 8.97$, $SE = 5.80$
Quadratic trend-by-Condition interaction	$F(1, 1086) = 0.69$, $p = .408$, $b = 0.78$, $SE = .946$
Model 2 (classic condition only)	
Random effects included	By-dyad intercepts; by-dyad slopes corresponding to the linear trend
Effect of linear trend	$F(1, 574) = 78.78$, $p < .001$, $b = -25.08$, $SE = 2.83$
Effect of quadratic trend	$F(1, 564) = 38.73$, $p < .001$, $b = 2.88$, $SE = 0.46$
Model 3 (new cards condition only)	
Random effects included	By-dyad and by-item intercepts; by-dyad slopes corresponding to Condition
Effect of linear trend	$F(1, 546) = 9.94$, $p = .002$, $b = -15.97$, $SE = 5.06$
Effect of quadratic trend	$F(1, 549) = 6.38$, $p = .012$, $b = 2.08$, $SE = 0.82$

Lexical diversity. Lexical diversity at Trial 2 was lower than the theoretical value of 1 in both conditions (classic: $b = -0.78$, $SE = 0.03$, $p < .001$; new cards: $b = -0.57$, $SE = 0.03$, $p < .001$). Moreover, lexical diversity at Trial 6 was lower than 1 in both conditions (classic: $b = -0.63$, $SE = 0.03$, $p < .001$; new cards: $b = -0.76$, $SE = 0.03$, $p < .001$). Results for Trials 2-5 are provided in Table 5. The significant negative linear trend for Trial in Model 1 shows that lexical diversity decreased in the classic condition. The significant positive quadratic trend shows that this decrease was less important in later trials in the classic condition. Neither the effect of Condition, nor the linear trend-by-Condition interaction or the quadratic trend-by-Condition interaction were significant. The effects of Trial were investigated further in Models 2 and 3. These revealed a significant negative linear trend and a significant positive quadratic trend in the classic condition, confirming the results from Model 1. They also revealed a significant negative linear trend and a significant positive quadratic trend in the new cards condition, also confirming the results from Model 1.

Table 5

Experiment 2 – Results of the Analysis on Lexical Diversity (Trials 2 – 5)

Model	**Model structure and coefficients**
Model 1 (All data)	
Random effects included	By-dyad and by-item intercepts; by-item slopes corresponding to Condition and the quadratic trend
Effect of linear trend	$F(1, 847) = 16.73$, $p < .001$, $b = -0.21$, $SE = 0.07$
Effect of quadratic trend	$F(1, 856) = 9.20$, $p = .003$, $b = 0.02$, $SE = 0.01$
Effect of Condition	$F(1, 873) = 2.01$, $p = .157$, $b = 0.25$, $SE = 0.17$
Linear trend-by-Condition interaction	$F(1, 846) = 0.04$, $p = .844$, $b = -0.02$, $SE = 0.11$
Quadratic trend-by-Condition interaction	$F(1, 851) < 0.01$, $p = .995$, $b > -0.01$, $SE = 0.02$
Model 2 (classic condition only)	
Random effects included	By-dyad intercepts; by-item slopes corresponding to the quadratic trend
Effect of linear trend	$F(1, 429) = 8.45$, $p = .004$, $b = -0.21$, $SE = 0.07$
Effect of quadratic trend	$F(1, 433) = 5.11$, $p = .024$, $b = 0.02$, $SE = 0.01$
Model 3 (new cards condition only)	
Random effects included	By-dyad and by-item intercepts; by-dyad slopes corresponding to the quadratic trend
Effect of linear trend	$F(1, 421) = 7.99$, $p = .005$, $b = -0.22$, $SE = 0.02$
Effect of quadratic trend	$F(1, 427) = 3.88$, $p = .049$, $b = 0.02$, $SE = 0.01$

The additional analysis revealed that lexical diversity at Trial 6 was lower in the new cards condition than in the classic condition, $F(1, 30) = 21.09$, $p < .001$. Lexical diversity at Trial 6 was higher than at Trial 5 in the classic condition, $b = -0.28$, $SE = 0.04$, $p < .001$, but not in the new cards condition, $b = -0.03$, $SE = 0.04$, $p = .498$.

Indefinite reference. Table 6 displays the results for Trials 1-5. No significant effects were found, regardless of whether we examined the entire dataset (Model 1) or both conditions separately (Models 2 and 3).

Table 6

Experiment 2 – Results of the Analysis on Indefinite Reference (Trials 1 to 5)

Model	**Model structure and coefficients**
Model 1 (All data)	
Random effects included	By-dyad intercepts; by-dyad slopes corresponding to the linear trend; by-item slopes corresponding to the quadratic trend
Effect of linear trend	$F(1, 1090) = 1.04$, $p = .308$, $b = - 0.03$, $SE = 0.01$
Effect of quadratic trend	$F(1\ 1149) = 0.55$, $p = .459$, $b < 0.01$, $SE < 0.01$
Effect of Condition	$F(1, 401) = 1.64$, $p = .202$, $b = - 0.03$, $SE = 0.02$
Linear trend-by-Condition interaction	$F(1, 1090) = 3.11$, $p = .078$, $b = 0.03$, $SE = 0.02$
Quadratic trend-by-Condition interaction	$F(1, 1134) = 1.55$, $p = .213$, $b > - 0.01$ (neg.), $SE < 0.01$
Model 2 (classic condition only)	
Random effects included	By-dyad intercepts; by-item slopes corresponding to the quadratic trend
Effect of linear trend	$F(1, 495) = 3.17$, $p = .076$, $b = - 0.03$, $SE = 0.01$
Effect of quadratic trend	$F(1, 562) = 1.80$, $p = .180$, $b < 0.01$, $SE < 0.01$
Model 3 (new cards condition only)	
Random effects included	By-dyad and by-item intercepts; by-dyad and by-item slopes corresponding to the linear trend
Effect of linear trend	$F(1, 535) = 0.40$, $p = .525$, $b = 0.01$, $SE = 0.01$
Effect of quadratic trend	$F(1, 551) = 0.19$, $p = .662$, $b > - 0.01$ (neg), $SE < 0.01$

The additional analysis revealed that indefinite reference production was similar in both conditions at Trial 6, $F(1, 28) = 3.53$, $p = .071$. In the classic condition, indefinite reference at Trial 6 was not significantly different from Trial 1, $b = -0.02$, $SE = 0.01$, $p = .146$. However, it was lower at Trial 5 than at Trial 6, $b = - 0.06$, $SE = 0.01$, $p < .001$. In the new cards condition, indefinite reference at Trial 6 was not different from Trial 1 or Trial 5, respectively $b = - 0.01$, $SE = 0.01$, $p = .291$ and $b > - 0.01$, $SE = 0.01$, $p = .835$.

Discussion

The additional test constituted by the introduction of Trial 6 with new cards in both conditions revealed a substantial and hitherto hidden benefit of the new cards condition. In terms of collaborative effort, new cards pairs were more efficient on Trial 6 than classic pairs were. Classic pairs had to suddenly use novel expressions that were not part of their common ground, as evidenced by an increase in lexical diversity and indefinite reference. Ultimately, classic pairs lost all benefits of lexical entrainment from their prior task completion, reverting to similar levels of collaborative effort as at Trial 1. These findings suggest that while referring to a larger set of referents does not allow achieving the high levels of coordination and efficiency afforded by repeated referring, it leads to more robust coordination, in the sense that participants are better able to adapt to novel referents.

There are at least three possible explanations for this finding, which are not mutually exclusive. First, new cards participants may have developed a greater familiarity with the idiosyncratic language use characteristics of their partners (Schober & Carstensen, 2010). Such idiosyncrasies may be more visible when partners talk about a wider range of referents than when they always talk about the same referents, or because of the longer negotiation of reference required by the ever-changing cards. Second, new cards participants may have developed an overarching meta-perspective on more generally recurring features of the tangram figures as surmised above. Third, and in parallel to the previous two explanations, the different ways of referring to targets in the two conditions may have fostered differences in what information participants attended to as relevant (E. Clark, 1997; Voiklis & Korter, 2012), but also in what features they remembered. Such differences might carry over to future referring situations when participants continue the task with a new partner who has no experience of it. In Experiment 3, we therefore compared new cards and classic directors who completed the task with a second matcher after having initially completed it with a first matcher.

In Trials 1-5, we also replicated the design of Experiment 1. Results are comparable to Experiment 1, but differ in two important ways. First, there is a decrease in collaborative effort over trials in both conditions (not only in the classic condition as in Experiment 1). Thus, new cards pairs were able to benefit from lexical entrainment to decrease collaborative effort (albeit less than classic pairs). Indeed, 11 of 15 new cards pairs decreased collaborative effort between Trial 1 and Trial 5 (compared with 14 of 15 classic pairs). We have thus demonstrated that being able to create conceptual pacts via repeated reference to the same targets is not necessary for improving efficiency in collaborative referring. Second, the rate of indefinite reference did not vary significantly over trials in the new cards condition (rather than increasing as it did in Experiment 1). This latter finding makes more sense theoretically and suggests that the increase found in Experiment 1 may have been a statistical anomaly.

Experiment 3

In Experiment 3, we ran both conditions (new cards and classic) again for four trials in order to test replicability of the findings in Experiments 1 and 2. Since much research on the matching task (e.g., H. Clark & Wilkes-Gibbs, 1986) shows that decreases in collaborative effort are greatest in earlier trials, we decided that four trials (Trials 1-4) would be enough to evidence similar phenomena as in Experiments 1 and 2. These four trials constituted Phase A of the experiment. In Phase B, directors from Phase A continued the task in the same way for another four trials (Trials 5-8), but with a new matcher who had not previously completed the task. That is, in the classic condition, directors discussed the same cards as in Phase A with the new matchers, whereas in the new cards condition, they discussed a different set of cards on each trial, just as in Phase A.

The classic condition in Experiment 3 is conceptually identical to one condition of a study performed by Wilkes-Gibbs and H. Clark (1992) (see also Gann & Barr, 2014). In that condition, directors completed the matching task with a first matcher (A) for six trials (Trials 1-6) and then continued the task with a second matcher (B) that had not done the task before, also for six trials (Trials 7-12). Pairs used more words to complete the task on Trial 7 than on Trial 6. However, they used fewer words on Trial 7 than on Trial 1. So there was a cost of changing matchers, but that cost was partly offset by some kind of gain. Because the matcher changed between the A and B trials, this cannot be due to common ground, and so the authors attributed the gain to the director's familiarity with the cards when doing the task the second time around. But the authors did not measure the evolution of lexical diversity. If directors in our study are familiar with the cards, then lexical diversity should not increase between Trial 4 and Trial 5 in our Experiment 3 (that is, the equivalent of Wilkes-Gibbs and H. Clark's Trials 6 and 7), because they would largely reuse the same descriptions they had entrained on with A matchers.

What will happen in the new cards condition? If collaborative effort decreases again during Phase A as it did in Experiment 2, will new cards participants in Phase B be affected by the introduction of a new matcher? If collaborative effort increases in Phase B, that would suggest that some or all of the benefits gained between directors and A matchers are not transferable to B matchers. If, however, directors are able to complete the task with B matchers without substantial increases in collaborative effort, that would be a further indication of the latent advantages afforded by repeatedly referring to new cards. That is, placing new cards on each trial may impact categorization processes of both partners (Markman & Makin, 1998; Voiklis & Korter, 2012), independently of the conceptual pacts or shared conversational history they may have elaborated together. This latter explanation would also be supported if lexical diversity were to remain at previous levels after the introduction of B matchers to the experiment.

Method

Participants. We recruited students from the student body of a Swiss university (N = 72, 50 women). They received 10 CHF for participating. Participants were randomly allocated to one of 24 groups of 3. Groups were randomly allocated to either the classic or new cards condition (12 groups in each condition). Roles (Director, Matcher A, Matcher B) were randomly allocated to participants.

Materials and procedure. The materials were similar to those used in Experiments 1 and 2. An additional sixteen cards were added to the pool of forty-eight cards from Experiment 2. For each group in the classic condition, a set of 8 cards was randomly drawn and used in each of 8 trials. For new cards groups, different cards were drawn without replacement for each of the 8 trials. The Director and Matcher A participated in the first 4 trials. Directors and A matchers were not informed that there would be a second phase in the experiment. In the B phase, which started immediately after the A phase, the Director and Matcher B participated in the next 4 trials.

Data preparation. Like in Experiments 1 and 2, we audio-recorded participants' talk and transcribed the recordings verbatim. Transcripts were then coded for each dependent variable (collaborative effort, lexical diversity and indefinite reference) per card and per trial following the same procedure as in Experiments 1 and 2.

Data analysis. For all dependent variables, we computed several models. Model 1 included Trial (linear and quadratic trends), Condition, Phase and their interaction terms as predictors. For collaborative effort and indefinite reference, Phase A was composed of Trials 1 to 4, and Phase B of Trials 5 to 8. For lexical diversity, Phase A was composed of Trials 2 to 4, and Phase B of Trials 6 to 8. Although lexical diversity in Trial 5 (i.e., the first Phase B trial) has variance, in order to analyze trial-trends by phase, it is necessary to have the same number of trials in Phase A and Phase B.

Model 2 included Trial (linear and quadratic trends), phase, and their interaction term for the classic condition. Model 3 included the same variables for the new cards condition. In addition, in Experiment 3 we ran additional models to test the effects of Phase. Model 4 thus included Trial (linear and quadratic trends) for Phase A of the classic condition, Model 5 for Phase A of the new cards condition, Model 6 for Phase B of the classic condition, and Model 7 for Phase B of the new cards condition. Additionally, for lexical diversity, the difference tests to a theoretical value were performed as in Experiments 1 and 2.

For all dependent variables, we also tested the difference between conditions at Trial 5. For this purpose, we computed a model that included Condition and Trial as factor (using Trial 5 as the reference category), as well as their interaction term. In that model, the effect of Condition corresponds to the difference between conditions at Trial 5.

We also tested the difference between Trials 5 and 4 for all three dependent variables, and between Trials 5 and 1 for collaborative effort and indefinite reference. We did so by computing additional two models (one per condition) that included Trial entered as a factor, using Trial 5 as the reference category. The estimates for Trials 1 and 4 represent, respectively, the difference between the value at Trial 5 and the value at Trial 1, and the difference between the value at Trial 5 and the value at Trial 4. Importantly, the models initially used to analyze lexical diversity did not allow us to test the difference between Trials 5 and 6 (this is because, as specified above, Phase A was composed of Trial 2 – 4 and Phase B was composed of Trial 6 – 8 in this analysis). These two additional models were thus also used to examine the difference between Trials 5 and 6 in more detail.

Results

Figure 3 presents the means by Trial and Condition (and standard errors of the means) for all three dependent variables in Experiment 3.

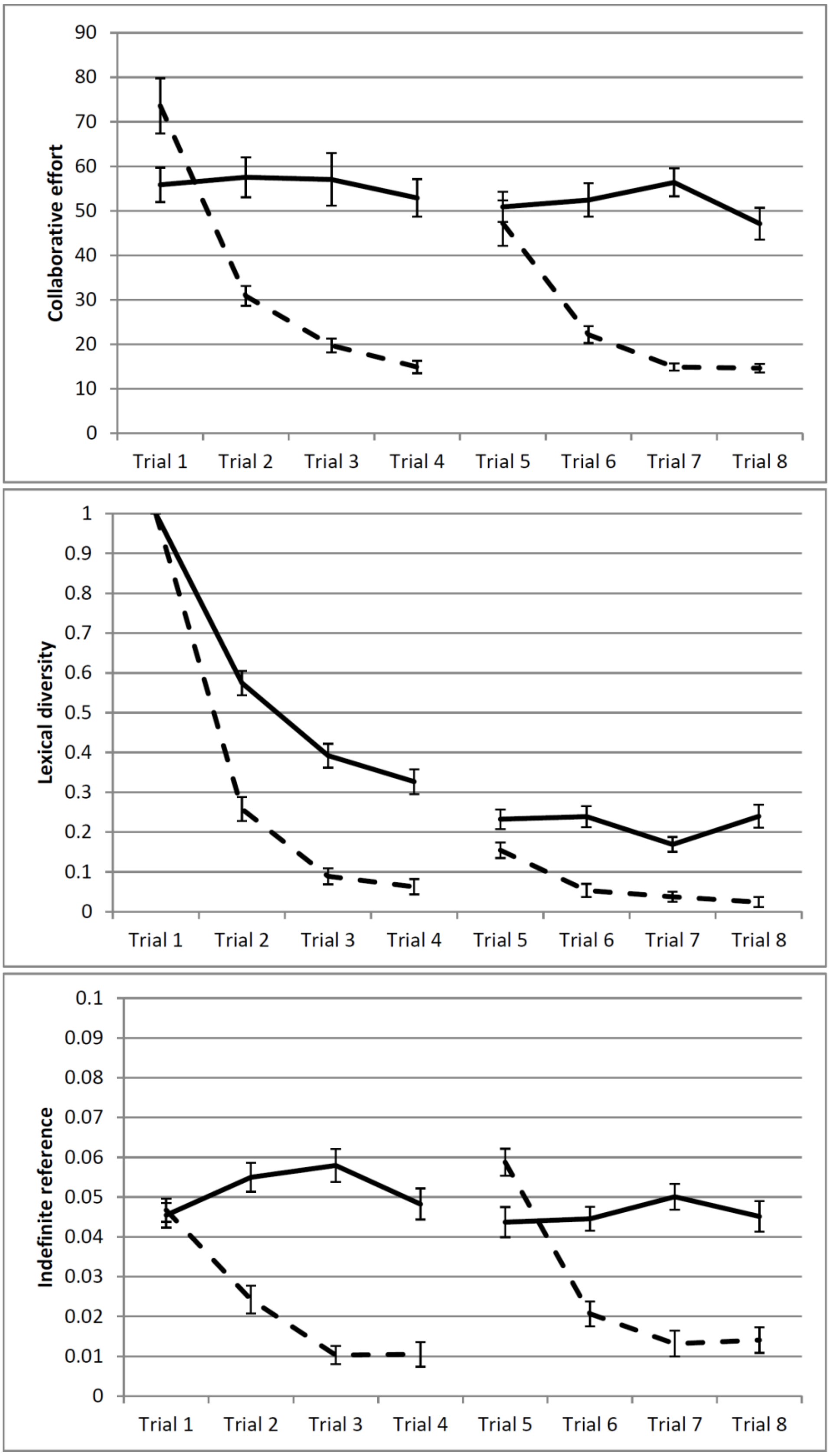

Figure 3: Collaborative effort (number of words per figure), lexical diversity (proportion of new lexical items divided by the total number of words used) and indefinite reference (number of indefinite

references divided by the total number of words used) by trial and condition for Experiment 3. Dashed lines represent the classic condition. Plain lines represent the new cards condition. Error bars represent the standard error of the mean.

Collaborative effort. Table 7 displays the results for Models 1 -7. The significant negative linear trend for Trial in Model 1 shows that collaborative effort decreased across trials in the classic condition. The significant positive quadratic trend shows that this decrease was less important in later trials in the classic condition. The significant effect of Condition shows that collaborative effort was smaller in the new cards condition than in the classic condition (controlling for Trial). We also found a significant Condition-by-Phase interaction, a significant linear trend-by-Condition interaction and a significant quadratic trend-by-Condition interaction. All other effects and interactions in the model failed to reach statistical significance. The effects of Trial were investigated further in Models 2 – 7.

Model 2 revealed a significant negative linear trend and a significant positive quadratic trend, confirming the results from Model 1 in the classic condition. There was also a significant effect of Phase, which shows that collaborative effort was greater in phase B than in phase A (controlling for Trial). The effects of Trial were investigated further in Models 4 and 6. Specifically, these two models confirmed that the negative linear trend and the positive quadratic trend were significant in both phase A and phase B of the classic condition.

All effects failed to reach statistical significance in Model 3, which focused on the data from the new cards condition. This was confirmed in Models 5 and 7, which confirmed that the linear and quadratic trends failed to reach statistical significance in both Phase A and Phase B of the new cards condition.

The additional analysis revealed that collaborative effort at Trial 5 was higher in the new cards condition than in the classic condition, $F(1, 85) = 27.23$, $p < .001$. In the classic condition, collaborative effort at Trial 5 was lower than at Trial 1, $b = 25.86$, $SE = 5.35$, $p < .001$, but higher than at Trial 4, $b = - 32.41$, $SE = 5.35$, $p < .001$. In the new cards condition, collaborative effort at Trial 5 was not significantly different from Trial 1 or Trial 4, respectively $b = 4.21$, $SE = 7.77$, $p = .590$ and $b = 2.28$, $SE = 7.76$, $p = .770$.

Table 7

Experiment 3 – Results of the Analysis on Collaborative Effort

Model	**Model structure and coefficients**
Model 1 (All data)	
Random effects included	By-dyad and by-item intercepts; by-dyad intercepts corresponding to Phase; by-item intercepts corresponding to Phase and Condition
Effect of linear trend	$F(1, 1303) = 14.01$, $p < .001$, $b = -66.07$, $SE = 8.51$
Effect of quadratic trend	$F(1, 1303) = 12.39$, $p < .001$, $b = 9.47$, $SE = 1.68$
Effect of condition	$F(1, 71) = 19.45$, $p < .001$, $b = -82.01$, $SE = 14.06$
Effect of phase	$F(1, 28) = 2.41$, $p = .131$, $b = 218.06$, $SE = 69.97$
Condition-by-phase interaction	$F(1, 1303) = 7.81$, $p = .005$, $b = -280.28$, $SE = 100.29$
Linear trend-by-Condition interaction	$F(1, 1303) = 31.44$, $p < .001$, $b = 76.19$, $SE = 12.16$
Linear trend-by-phase interaction	$F(1, 1303) = 0.17$, $p = .679$, $b = -24.97$, $SE = 23.41$
Linear trend-by-condition by phase interaction	$F(1, 1303) = 1.16$, $p = .283$, $b = 36.04$, $SE = 33.53$
Quadratic trend-by-Condition interaction	$F(1, 1303) = 32.56$, $p < .001$, $b = -11.56$, $SE = 2.39$
Quadratic trend-by-phase interaction	$F(1, 1303) = 0.69$, $p = .407$, $b = -3.28$, $SE = 2.37$
Quadratic trend-by-condition by phase interaction	$F(1, 1303) = 1.22$, $p = .269$, $b = 3.74$, $SE = 3.39$
Model 2 (classic condition only)	
Random effects included	By-dyad and by-item intercepts
Effect of linear trend	$F(1, 702) = 59.37$, $p < .001$, $b = -66.04$, $SE = 7.41$
Effect of quadratic trend	$F(1, 702) = 57.66$, $p < .001$, $b = 9.47$, $SE = 1.46$
Effect of phase	$F(1, 702) = 12.82$, $p < .001$, $b = 218.07$, $SE = 60.91$
Linear trend-by-phase interaction	$F(1, 702) = 1.50$, $p = .221$, $b = -24.99$, $SE = 20.39$
Quadratic trend-by-phase interaction	$F(1, 702) = 2.51$, $p = .113$, $b = -3.27$, $SE = 2.06$
Model 3 (new cards condition only)	
Random effects included	By-dyad and by-item intercepts; by-dyad and by-item slopes corresponding to Phase
Effect of linear trend	$F(1, 612) = 1.63$, $p = .202$, $b = 10.07$, $SE = 9.66$
Effect of quadratic trend	$F(1, 612) = 2.13$, $p = .145$, $b = -2.09$, $SE = 1.90$
Effect of phase	$F(1, 14) = 0.80$, $p = .387$, $b = -71.59$, $SE = 80.21$
Linear trend-by-phase interaction	$F(1, 612) = 0.28$, $p = .597$, $b = 14.17$, $SE = 26.79$
Quadratic trend-by-phase interaction	$F(1, 612) = 0.01$, $p = .929$, $b = 0.24$, $SE = 2.70$
Model 4 (classic condition, Phase A)	
Random effects included	By-dyad and by-item intercepts
Effect of linear trend	$F(1, 321) = 63.46$, $p < .001$, $b = -66.05$, $SE = 8.29$
Effect of quadratic trend	$F(1, 321) = 33.62$, $p < .001$, $b = 9.46$, $SE = 1.63$
Model 5 (new cards condition, Phase A)	
Random effects included	By-dyad intercepts; by-dyad and by-item slopes corresponding to the linear trend
Effect of linear trend	$F(1, 10) = 0.63$, $p = .447$, $b = 8.26$, $SE = 10.43$
Effect of quadratic trend	$F(1, 295) = 0.77$, $p = .382$, $b = -1.780$, $SE = 2.05$
Model 6 (classic condition, Phase B)	
Random effects included	By-dyad and by-item intercepts
Effect of linear trend	$F(1, 322) = 29.07$, $p < .001$, $b = -91.03$, $SE = 16.89$
Effect of quadratic trend	$F(1, 322) = 22.86$, $p < .001$, $b = 6.19$, $SE = 1.30$
Model 7 (new cards condition, Phase B)	
Random effects included	By-dyad and by-item intercepts
Effect of linear trend	$F(1, 10) = 1.99$, $p = .189$, $b = 31.19$, $SE = 22.12$
Effect of quadratic trend	$F(1, 295) = 2.05$, $p = .153$, $b = -2.43$, $SE = 1.70$

Lexical diversity. Lexical diversity at Trial 2 was lower than the theoretical value of 1 in both the classic condition and the new cards condition, respectively b = - 0.74, SE = 0.02, p < .001 and b = - 0.43, SE = 0.03, p < .001. Lexical diversity at Trial 5 was also lower than the theoretical value of 1 in both the classic condition and the new cards condition, respectively b = - 0.85, SE = 0.02, p < .001 and b = - 0.76, SE = 0.03, p < .001.

Results for Trials 2 – 4 and 6 – 8 are displayed in Table 8. The significant negative linear trend for Trial in Model 1 shows that lexical diversity decreased across trials in the classic condition. The significant positive quadratic trend shows that this decrease was less important in later trials in the classic condition. All other effects and interactions in the model failed to reach statistical significance. The effects of Trial were investigated further in Models 2-7.

Model 2 revealed a significant positive quadratic trend in the classic condition, partly confirming the results from Model 1. There was also a significant quadratic trend-by-Phase interaction, which was analyzed further in Models 4 and 6. Model 4 revealed that there was a significant negative linear trend and a significant positive linear trend in phase A of the classic condition; however, Model 6 revealed that these effects were no longer significant in phase B.

Model 3 revealed a significant negative linear trend and a significant positive quadratic trend in the new cards condition, confirming the results from Model 1. These findings were analyzed further in Models 5 and 7. Model 5 revealed that only the negative linear trend was significant in phase A of the new cards condition; Model 7 revealed that this was also the case in phase B.

Lexical diversity at Trial 5 was higher in the new cards condition than in the classic condition, F(1, 96) = 87.41, p < .001. In the classic condition, lexical diversity at Trial 5 was higher than lexical diversity at Trial 4 and at Trial 6, respectively b = - 0.09, SE = 0.03, p = .002 and b = - 0.10, SE = 0.03, p = 0.01. In the new cards condition, lexical diversity at Trial 5 was lower than lexical diversity at Trial 4, but not significantly different from lexical diversity at Trial 6, respectively b = 0.10, SE = 0.04, p = .031 and b = - 0.01, SE = 0.04, p = .808.

Table 8

Experiment 3 – Results of the Analysis on Lexical Diversity (Trials 2 - 4 and 5- 8)

Model	**Model structure and coefficients**
Model 1 (All data)	
Random effects included	By-dyad intercepts; by-dyad slopes corresponding to Phase; by-item slopes corresponding to Condition and Phase
Effect of linear trend	$F(1, 905) = 9.66$, $p = .002$, $b = - 0.54$, $SE = 0.17$
Effect of quadratic trend	$F(1, 905) = 12.22$, $p = .001$, $b = 0.07$, $SE = 0.03$
Effect of condition	$F(1, 90) = 2.55$, $p = .114$, $b = 0.29$, $SE = 0.35$
Effect of phase	$F(1, 28) = 0.19$, $p = .663$, $b = - 0.85$, $SE = 1.37$
Condition-by-phase interaction	$F(1, 905) = 1.70$, $p = .192$, $b = 2.57$, $SE = 1.97$
Linear trend-by-Condition interaction	$F(1, 905) = 1.59$, $p = .207$, $b = 0.02$, $SE = 0.24$
Linear trend-by-phase interaction	$F(1, 905) = 0.10$, $p = .747$,$b = 0.51$, $SE = 0.43$
Linear trend-by-Condition-by-phase interaction	$F(1, 905) = 1.79$, $p = .182$, $b = - 0.82$, $SE = 0.61$
Quadratic trend-by-Condition interaction	$F(1, 905) = 0.81$, $p = .368$, $b = - 0.01$, $SE = 0.04$
Quadratic trend-by-phase interaction	$F(1, 905) = 1.93$, $p = .165$, $b = - 0.07$, $SE = 0.04$
Quadratic trend-by-condition by phase interaction	$F(1, 905) = 1.28$, $p = .258$, $b = 0.06$, $SE = 0.06$
Model 2 (classic condition only)	
Random effects included	By-item intercepts; by-dyad and by-item slopes corresponding to Phase
Effect of linear trend	$F(1, 448) = 2.60$, $p = .107$, $b = - 0.54$, $SE = 0.14$
Effect of quadratic trend	$F(1, 448) = 5.15$, $p = .024$, $b = 0.07$, $SE = 0.02$
Effect of phase	$F(1, 22) = 0.57$, $p = .459$, $b = - 0.85$, $SE = 1.13$
Linear trend-by-phase interaction	$F(1, 448) = 2.11$, $p = .147$, $b = 0.51$, $SE = 0.35$
Quadratic trend–by-phase interaction	$F(1, 448) = 4.88$, $p = .028$, $b = - 0.07$, $SE = 0.03$
Model 3 (new cards condition only)	
Random effects included	By-dyad and by-item intercepts; by-dyad and by-item slopes corresponding to Phase
Effect of linear trend	$F(1, 409) = 6.97$, $p = .009$, $b = - 0.52$, $SE = 0.20$
Effect of quadratic trend	$F(1, 409) = 7.16$, $p = .008$, $b = 0.07$, $SE = 0.03$
Effect of phase	$F(1, 14) = 1.08$, $p = .317$, $b = 1.70$, $SE = 1.64$
Linear trend-by-phase interaction	$F(1, 409) = 0.35$, $p = .556$, $b = - 0.30$, $SE = 0.51$
Quadratic trend-by-phase interaction	$F(1, 409) = 0.04$, $p = .849$, $b = - 0.01$, $SE = 0.05$
Model 4 (classic condition, Phase A)	
Random effects included	By-item intercepts
Effect of linear trend	$F(1, 233) = 10.15$, $p = .002$, $b = - 54.11$, $SE = 0.17$
Effect of quadratic trend	$F(1, 233) = 6.85$, $p = .010$, $b = 0.07$, $SE = 0.03$
Model 5 (new cards condition, Phase A)	
Random effects included	By-dyad and by-item intercepts; by-dyad and by-item slopes corresponding to the quadratic trend
Effect of linear trend	$F(1, 136) = 5.93$, $p = .016$, $b = - 0.54$, $SE = 0.22$
Effect of quadratic trend	$F(1, 10) = 3.61$, $p = .087$, $b = 0.07$, $SE = 0.04$
Model 6 (classic condition, Phase B)	
Random effects included	By-dyad and by-item intercepts; by-dyad and by-item slopes corresponding to the linear trend and the quadratic trend
Effect of linear trend	$F(1, 226) = 0.01$, $p = .904$, $b = - 0.03$, $SE = 0.24$
Effect of quadratic trend	$F(1, 226) < 0.01$, $p = .953$, $b < 0.01$, $SE = 0.02$
Model 7 (new cards condition, Phase B)	
Random effects included	By-dyad and by-item intercepts; by-item slopes corresponding to the quadratic trend
Effect of linear trend	$F(1, 149) = 3.96$, $p = .049$, $b = - 0.83$, $SE = 0.42$
Effect of quadratic trend	$F(1, 60) = 3.95$, $p = .052$, $b = 0.06$, $SE = 0.03$

Indefinite reference. Table 9 displays the results for Models 1 - 7. The significant negative linear trend for Trial in Model 1 shows that indefinite reference production decreased across trials in the classic condition. The significant positive quadratic trend shows that this decrease was less important in later trials in the classic Condition. The significant effect of Condition reveals that indefinite reference production was smaller in the new cards condition than in the classic condition (controlling for Trial). The significant effect of Phase reveals that indefinite reference production was greater in phase B than in phase A (also controlling for Trial). What is more, all interactions in the model were significant, except for the quadratic trend by Condition by Phase interaction. The effects of Trial were investigated further in Models 2 – 7.

All effects and interactions in Model 2 were significant, confirming the results from Model 1 in the classic condition. These findings were analyzed further in Models 4 and 6. These models revealed a significant negative linear trend and a significant positive quadratic trend in both phases of the classic condition.

All effects and interactions in Model 3 failed to reach statistical significance, suggesting that the results from Model 1 could not be generalized to the new cards condition. These findings were analyzed further in Models 5 and 7. Model 5 revealed a significant *positive* linear trend and a *negative* quadratic trend in phase A of the new cards condition. Model 7 revealed that neither trend was significant in phase B.

The additional analysis revealed that indefinite reference production at Trial 5 was higher in the classic condition than in the new cards condition, $F(1, 85) = 30.84$, $p < .001$. In the classic condition, indefinite reference production at Trial 5 was higher than at Trial 1 and at Trial 4, respectively $b = -0.01$, $SE = 0.01$, $p = .011$ and $b = -0.05$, $SE = 0.01$, $p < .001$. In the new cards condition, indefinite reference production at Trial 5 was not significantly different from Trial 1 or Trial 4, respectively $b < 0.01$, $SE = 0.01$, $p = .621$ and $b < 0.01$, $SE = 0.01$, $p = .921$.

Table 9

Experiment 3 – Results of the Analysis on Indefinite Reference

Model	**Model structure and coefficients**
Model 1 (All data)	
Random effects included	By-dyad and by-item intercepts; by-dyad slopes corresponding to Phase and by-item slopes corresponding to Phase and Condition
Effect of linear trend	$F(1, 1303) = 21.46$, $p < .001$, $b = - 0.04$, $SE = 0.01$
Effect of quadratic trend	$F(1, 1303) = 9.60$, $p = .002$, $b = 0.01$, $SE < 0.01$
Effect of condition	$F(1, 71) = 38.76$, $p < .001$, $b = - 0.06$, $SE = 0.06$
Effect of phase	$F(1, 28) = 19.73$, $p < .001$, $b = 0.44$, $SE = 0.09$
Condition-by-phase interaction	$F(1, 1303) = 25.32$, $p < .001$, $b = 0.07$, $SE = 0.01$
Linear trend-by-Condition interaction	$F(1, 1303) = 50.78$, $p < .001$, $b = 0.07$, $SE = 0.01$
Linear trend-by-phase interaction	$F(1, 1303) = 13.00$, $p < .001$, $b = - 0.10$, $SE = 0.03$
Linear trend-by-condition-by-phase interaction	$F(1, 1303) = 8.49$, $p = .004$, $b = 0.09$, $SE = 0.03$
Quadratic trend-by-Condition interaction	$F(1, 1303) = 46.05$, $p < .001$, $b = - 0.01$, $SE < 0.01$
Quadratic trend-by-phase interaction	$F(1, 1303) = 6.34$, $p = .012$, $b < 0.01$, $SE < 0.01$
Quadratic trend-by-condition-by-phase interaction	$F(1, 1303) = 0.01$, $p = .908$, $b > - 0.01$ (neg.), $SE < 0.01$
Model 2 (classic condition only)	
Random effects included	By-dyad and by-item intercepts; by-dyad slopes corresponding to Phase and the linear trend; by-item slopes corresponding to the linear trend
Effect of linear trend	$F(1, 6) = 80.27$, $p < .001$, $b = - 0.04$, $SE = 0.01$
Effect of quadratic trend	$F(1, 632) = 56.66$, $p < .001$, $b = 0.01$, $SE < 0.01$
Effect of phase	$F(1, 18) = 52.09$, $p < .001$, $b = 0.44$, $SE = 0.06$
Linear trend-by-phase interaction	$F(1, 632) = 24.57$, $p < .001$, $b = - 0.10$, $SE = 0.02$
Quadratic trend-by-phase interaction	$F(1, 632) = 3.98$, $p = .046$, $b < 0.01$, $SE < 0.01$
Model 3 (new cards condition only)	
Random effects included	By-dyad and by-item intercepts; by-dyad slopes corresponding to Phase and the quadratic trend; by-item slopes corresponding to Phase
Effect of linear trend	$F(1, 601) = 2.68$, $p = .102$, $b = 0.02$, $SE = .001$
Effect of quadratic trend	$F(1, 6) = 5.83$, $p = .052$, $b > - 0.01$ (neg.), $SE < 0.01$
Effect of phase	$F(1, 18) = 0.16$, $p = .696$, $b = - 0.03$, $SE = 0.07$
Linear trend-by-phase interaction	$F(1, 601) = 0.19$, $p = .659$, $b = - 0.01$, $SE = 0.02$
Quadratic trend-by-phase interaction	$F(1, 601) = 2.44$, $p = .119$, $b < 0.01$, $SE < 0.01$
Model 4 (classic condition, Phase A)	
Random effects included	By-dyad and by-item intercepts; by-dyad slopes corresponding to the linear trend
Effect of linear trend	$F(1, 10) = 32.46$, $p < .001$, $b = - 0.04$, $SE = 0.01$
Effect of quadratic trend	$F(1, 310) = 16.49$, $p < .001$, $b = 0.01$, $SE < 0.01$
Model 5 (new cards condition, Phase A)	
Random effects included	By-dyad intercepts; by-dyad and by-item slopes corresponding to the linear trend
Effect of linear trend	$F(1, 10) = 9.47$, $p = .012$, $b = 0.03$, $SE = 0.01$
Effect of quadratic trend	$F(1, 295) = 8.81$, $p = .003$, $b = - 0.01$, $SE < 0.01$
Model 6 (classic condition, Phase B)	
Random effects included	By-dyad and by-item intercepts; by-item slopes corresponding to the linear trend
Effect of linear trend	$F(1, 47) = 51.66$, $p < .001$, $b = - 0.14$, $SE = 0.02$
Effect of quadratic trend	$F(1, 274) = 42.00$, $p < .001$, $b = 0.01$, $SE < 0.01$
Model 7 (new cards condition, Phase B)	
Random effects included	By-dyad and by-item intercepts
Effect of linear trend	$F(1, 306) = 0.42$, $p = .516$, $b = 0.01$, $SE = 0.02$
Effect of quadratic trend	$F(1, 306) = 0.38$, $p = .540$, $b > - 0.01$ (neg.), $SE < 0.01$

Discussion

In Experiment 3, Trials 1-4 (Phase A) replicated the design of Experiments 1 and 2. Unlike Experiment 2, but similar to Experiment 1, we did not find a significant decrease in collaborative effort in the new cards condition (6 pairs out of 12 had a lower level of collaborative effort at Trial 4 of Phase A, compared with 12 out of 12 classic pairs). However, lexical diversity decreased in the new cards condition in much the same way as in Experiments 1 and 2. We found an inverted u-shaped curve for indefinite reference in the new cards condition, unlike Experiment 1 (where we found an increase over trials) and Experiment 2 (where we found no significant effects).

In Phase B, we tested the effects of a new matcher. There was little evidence that new cards pairs were affected by the change of matchers. Collaborative effort did not increase significantly, lexical diversity decreased in what seems to be an asymptotic manner over both phases, and indefinite reference did not significantly decrease. On the other hand, classic pairs evidenced a substantial (but transient) increase in collaborative effort (a finding that replicates Wilkes-Gibbs and H. Clark, 1992), a slight and transient increase in lexical diversity, and a substantial but transient increase in indefinite reference. In other words, new cards pairs were able to better adapt a new matcher at Trial 5, but classic pairs quickly reached the level of performance they had attained with the original matcher again.

These results are similar to those of Experiment 2 in some respects. That new cards pairs were little affected by the change of matchers resembles the finding from Experiment 2 that the introduction of new cards was easier for new cards pairs to cope with relative to classic pairs. On the other hand, evidence for new cards pairs' ability to decrease collaborative effort is inconsistent, with a significant decrease only found in Experiment 2 but not in Experiments 1 and 3. In the General Discussion, we now summarize the overall findings of Experiments 1-3 and draw out implications.

General Discussion

When people repeatedly interact with each other, they come to use the same words to describe recurrent objects of reference. This phenomenon, known as lexical entrainment, is robust (Brennan & H. Clark, 1996; Garrod & Anderson, 1987). Lexical entrainment has often been identified with the establishment of conceptual pacts, a process by which conversational partners agree to use certain terms of reference with each other to refer to specific objects (Brennan & H. Clark, 1996). However, lexical entrainment may occur as partners develop a shared conversational history, potentially facilitating future conversations even when partners are not able to develop and reuse conceptual pacts,

We thus conducted three experiments to systematically unconfound these two potential effects, by contrasting two conditions (classic and new cards). In the new cards condition, participants were unable to reuse any conceptual pacts they may have established on a given trial. We investigated collaborative effort, lexical diversity and indefinite reference as indicators of increased efficiency, lexical entrainment and the use of conceptual pacts, respectively. Experiment 1 showed that new cards pairs decreased lexical diversity over 5 trials, indicating that lexical entrainment occurred in the absence of conceptual pacts (as evidenced by the higher rates of indefinite reference). However, new cards pairs did not significantly decrease their collaborative effort. In Experiment 2, we attempted to replicate Experiment 1 and added a sixth trial where pairs in both conditions placed a set of new cards. We found that, again, new cards pairs showed evidence of lexical entrainment without conceptual pacts. Moreover, new cards decreased collaborative effort. On the sixth trial, classic pairs evidenced an increase in collaborative effort, lexical diversity and indefinite reference, as they were suddenly forced to renegotiate conceptual pacts to complete the task, suggesting that they were less able to accommodate to novel stimuli than new cards pairs. In Experiment 3, we attempted another replication in four initial trials (Phase A), but further had directors complete the task with a naïve matcher for four additional trials (Phase B). In Phase A, collaborative effort decreased over trials in classic pairs but not in new cards pairs, similarly to Experiment 1 but unlike Experiment 2. Lexical diversity decreased over trials in both conditions, replicating Experiments 1 and 2. Indefinite reference decreased over trials for classic pairs but not for new cards pairs, in a more or less similar fashion to Experiments 1 and 2. In other words, in all experiments, classic pairs were increasingly using conceptual pacts but new cards pairs were not. In Phase B, the introduction of a new matcher was more perturbing for classic pairs than new cards pairs, in that collaborative effort increased strongly from Trial 4 to Trial 5. Participants in the new cards condition found it easier not only to establish new terms, but also to establish new terms with new partners.

Lexical entrainment without conceptual pacts?

The most robust finding emerging from Experiments 1-3 is the decrease in lexical diversity that was substantial in both conditions, albeit more so in the classic condition than in the new cards condition. What is the cause of the lexical entrainment effect in the new cards condition? One potential explanation is that new cards pairs implicitly entrained on a “meta-perspective” that was shared by all the cards they referred to, i.e., that all cards were humanoid figures with corresponding features like heads, limbs and the like. This meta-perspective may help new cards pairs converge on more efficient procedural routines (Mills, 2013) for describing novel stimuli. For example, such a routine might contain slots for describing figures’ heads, limbs or bodies. Such templates might reflect new cards pairs’ experience with the specific attributes that are likely to remain constant over the stimuli (e.g., all figures have “heads”), while focusing their attention on variations within those attributes (e.g., a new figure’s “head” is shaped somewhat differently), as suggested by Markman and Makin (1998).

The meta-perspective explanation is consistent with the fact that classic pairs experienced more difficulties to adapt to novel stimuli than new cards pairs. In Experiment 2, classic pairs who suddenly had to deal with new cards at Trial 6 evidenced an increase in collaborative effort. Their previously elaborated conceptual pacts were useless to them (they reverted to levels of collaborative effort similar to Trial 1). In Experiment 3, classic pairs who changed matchers for the B phase of the experiment experienced similar issues. However, the account by which new cards pairs might, by discussing a wider range of referents, learn more about their partners’ conversational idiosyncrasies, seems to be nunanced by Experiment 3 findings: The adaptability benefit of new cards pairs extended to the situation with new matchers, suggesting that it is at least partly independent of the interactions with the initial matcher. Indeed, descriptive precedents entrained on by directors and initial matchers might be reused by directors in their subsequent interactions with new matchers. Further studies which vary the similarities between new cards might investigate this possibility. For example, contrasting a condition with new cards that share an overarching category membership with a condition with new cards that do not would allow testing the meta-perspective explanation.

To further investigate the meta-perspective explanation, we conducted supplementary analyses of the content of participants’ utterances in Experiments 1-3. These analyses focused on (1) the number of body parts (e.g., *head*, *arm*) and (2) the number of geometrical shapes (e.g., *triangle*, *square*) mentioned in each description (see Supplementary Analysis for details of methods and results). Results show that in Experiments 1 and 3, mentions of body parts and geometrical shapes decrease over trials in the classic condition, but not in the new cards condition, while in Experiment 2 mentions of body parts and geometrical shapes decrease over trials in both conditions. Overall, these results support the explanation that participants in the new cards condition develop an overarching meta-perspective on the task in their lexical choices, because they continue to use these lexical items over trials. Experiment 2 does not follow this pattern, but the decrease in body parts and geometrical shapes may parallel the significant decrease in collaborative effort found in that experiment.

The question arises how the concept of an "overarching meta-perspective" is best characterized. How is it similar to or different from a conceptual pact? Can such a meta-perspective be considered analogous to a conceptual pact? There are three reasons why we believe it cannot. First, what is typically understood as a conceptual pact in the literature (i.e., a partner-specific agreement about how to refer to a specific target; Brennan & Clark, 1996) only emerges progressively over repeated trials. Because participants typically initially discuss multiple features of figures in order to gain certainty that they are talking about the same figure, it is not always predictable from initial discussions which of those aspects will emerge as part of the "final" version of the pact (in fact, during Trial 1, participants do not even know they will be doing the same task again, so an agreement about how to refer to a figure in the future may not even be on their minds). As an example, in Experiment 1, a classic pair director referred to a figure on Trial 1 using four descriptors (our translations from French): *riding a bike, lying in bed, small* and *got a leg up*. In Trial 2, *riding a bike* and *lying down* were reused. In Trial 3, *riding a bike* was reused. In Trial 4, the director switched to *cyclist*, reusing this label again on Trial 5. This example illustrates how substantial variation in the choice of referring expressions persists over trials. Conceptual variation persisted until Trial 3, where the concept of a *bike rider* stabilized. Lexical variation persisted until Trial 4 (fluctuating between *riding a bike* and *cyclist*). Second, meta-level labels like geometrical shapes or body parts will be specific to a given figure and have to be adapted to any new referent. For example, what counts as an "arm" or a "leg" in a new figure will not be exactly the same for each new figure. So participants have to re-negotiate use of these terms. Third, participants themselves do not seem to treat references featuring body parts or geometric shapes as pacts. A conceptual pact has a normative element, that is, participants expect them to be used once established. Thus, they are marked as constituting shared knowledge, for example in the use of definite reference. However, the data consistently show that the level of indefinite reference remains high over trials in the new cards condition.

What participants in the classic matching task do (documented by the extensive qualitative analyses in Clark & Wilkes-Gibbs, 1986) is gradually converge on and consolidate an agreed-upon label. Initial descriptions can be seen as *precedents*. Over trials, some of these precedents will be reused and thus strengthened, while others will be abandoned. While it is difficult to predict which descriptions participants will converge on, this process may be influenced by several factors, for example the memory constraints of the participants (as suggested in Gann & Barr, 2014 or Knutsen & Le Bigot, 2014, where participants tend to reuse descriptions they themselves produce). A conceptual pact should thus be seen as the gradually emerging result of a process of negotiation and convergence (and thus conventionalization), rather than as an explicit agreement that is stabilized at the end of Trial 1. Part of the reason conceptual pacts may have been treated in the literature as more “stable” than they actually are is the term “pact” itself, which suggests an explicit and formally ratified agreement. While matching task participants sometimes do make such explicit agreements (e.g., *let’s call this one the cyclist*), it is more frequent for them to emerge implicitly, as a result of accumulation of precedents.

According to Brennan (2005), convergence on a conceptual pact is a process of reciprocal hypothesis production and testing on the part of the director and addressee. The director’s utterance constitutes a hypothesis on what the matcher might recognize as being the target figure, whereas the addressee entertains hypotheses about which figure is the target. This process moves from initial precedents to stable conceptual pacts. Based on the findings from Experiments 1-3, we suggest its efficiency depends on 2 factors. The first, corresponding roughly to the presentation phase in Clark and Schaefer’s (1989) contribution model, is the availability of descriptions. Figures that are easy to describe (because they correspond to widely shared labels, like “triangle”, or because they have been described before and are familiar) yield more adapted descriptions more quickly, requiring less words. The second factor, which corresponds roughly to the acceptance phase in Clark and Schaefer (1989), is the ease with which uncertainty about hypotheses can be reduced. Figures that have been described before multiple times feature very low uncertainty about their identity. In terms of these two factors, at the beginning of the classical matching task (with unacquainted participants), the availability of descriptions is low and uncertainty about possible target figure identities based on those descriptions is high. At the end of several trials in the classic condition, the availability of multiple descriptions is a bit higher and the uncertainty about how or which descriptions apply to target figures is very low (resulting in the signature patterns of low collaborative effort, low lexical diversity and more definite reference). At the end of several trials in the new cards condition, multiple descriptions are available but the uncertainty about how to apply those descriptions remains rather high (resulting in the signature patterns of low lexical diversity and more indefinite reference, with collaborative effort varying depending on how successful participants are in managing uncertainty).

Taken together, these findings suggest that, even when confronted with recurring novel referents, pairs who interact together are able to create some kind of entrainment. This is the direct and isolated effect of a shared interactional history, and is distinct from the effect of repeatedly referring to the same referents. Thus, this effect is also separate from the establishment of conceptual pacts, as evidenced by the lack of decrease in indefinite reference. In other words, our experiments show that, in the matching task, there is a “pure” effect of interacting together on lexical entrainment, an effect that is independent of the effect of conceptual pacts; however, this effect is not always strong enough to enable a systematic reduction in collaborative effort. Further research could investigate what conditions might enable pairs to reduce collaborative effort systematically.

To conclude, lexical entrainment in the matching task may be due to a range of other factors than the establishment of conceptual pacts (we do not mean to suggest conceptual pacts are unimportant, but rather that they are not the only phenomenon responsible for the decrease in collaborative effort). In our experiments, we investigated factors related to the interaction per se, including getting to know one’s partner’s idiosyncrasies (an explanation excluded by the results of Experiment 3) or the development of a broader meta-perspective through referring to a wider range of similar targets (an explanation supported by various lines of evidence, including the supplementary analysis). But other factors potentially contributing to lexical entrainment (or even directly to reduced collaborative effort) may be related to procedural coordination (Knutsen et al., 2019; Mills, 2013): Repeatedly doing a task may establish tacit routines that reduce the need for explicitly coordinating on the task. Further research should seek to establish the main causes (alongside conceptual pacts) of lexical entrainment and the reduction of collaborative effort in the matching task in order to understand better this crucial task for studying collaborative referring.

Acknowledgments

We thank Janet Bavelas for interesting discussions on Experiment 1 and Myriam Aribot, Nicolas Brandt, Charlotte Cavin and Charlotte Hart for assistance in conducting the experiments and coding data.

References

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science, 27*, 195-225.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Bortfeld, H., & Brennan, S. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes, 23*, 21-49.

Boyle, E. A., Anderson, A. H., & Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem-solving task. *Language and Speech, 37*, 1-20.

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions, 95-129.

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

Clark, E. V. (1997) Conceptual perspective and lexical choice in acquisition. *Cognition*, 64, 1-37.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Murphy, G. L. (1983). Audience design in meaning and reference. In J. F. LeNy and W. Kintsch (Eds.), Language and comprehension (pp. 287-299). Amsterdam: North-Holland Publishing Co.

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

De Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science, 4*, 232-248.

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive Science, 40*(1), 145-171.

Fussell, S. R., & Krauss, R. M. (1989). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology, 19,* 509-526.

Gann, T. M., & Barr, D. J. (2014). Speaking from experience: Audience design as expert performance. *Language, Cognition and Neuroscience, 29*(6), 744-760.

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

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (Vol. 3, pp. 41–58). New York: Academic Press.

Holler, J., Kendrick, K. H., Casillas, M., & Levinson, S. C. (Eds.) (2016). *Turn-taking in human communicative interaction*. Lausanne: Frontiers Media.

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

Hupet, M., Seron, X., & Chatraine, Y. (1991). The effect of the codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology, 82,* 449-462.

Isaacs, E. A., & Clark, H. H. (1987). References in conversations between experts and novices. *Journal of Experimental Psychology: General, 116,* 26-37.

Keysar, B. (1997). Unconfounding common ground. *Discourse Processes, 24,* 253-270.

Knutsen, D., Bangerter, A., & Mayor, E. (2019). Procedural coordination in the matching task. *Collabra: Psychology, 5,* 3. doi:10.1525/collabra.188

Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin and Review, 21,* 1590-1599. doi:10.3758/s13423-014-0620-7

Knutsen, D., Ros, C., & Le Bigot, L. (2018). Spoilt for choice: Initially considering several referential expressions affects subsequent referential decisions. *Language, Cognition and Neuroscience*, *33,* 618-632, doi: 10.1080/23273798.2017.1400080

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343-346.

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General, 127*, 331-354.

Mills, G. J. (2014) Dialogue in joint activity: complementarity, convergence, conventionalization. *New Ideas in Psychology, 32*, 158-173.

Richards, B. (1987). Type/Token Ratios: what do they really tell us? *Journal of Child Language, 14*, 201-209.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language, 50*, 696-735.

Schober, M.F., & Carstensen, L.L. (2010). Does being together for years help comprehension? In E. Morsella (Ed.), *Expressing oneself/Expressing one's self: Communication, cognition, language, and identity* (pp. 107-124). New York: Taylor & Francis.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*, 211-232.

Tolins, J., Zeamer, C., & Fox Tree, J. E. (2017). Overhearing dialogues and monologues. How does entrainment lead to more comprehensible referring expressions? *Discourse Processes*, advanced online publication.

Van der Wege, M. M. (2009). Lexical entrainment and lexical differentiation in reference phrase choice. *Journal of Memory and Language, 60*, 448-463.

Voiklis, J., & Corter, J. E. (2012). Conventional wisdom: Negotiating conventions of reference enhances category discovery. *Cognitive Science, 36*, 607-634.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 31*, 183-194.