THE UNIVERSITY *of* EDINBURGH

Interaction promotes the adaptation of referential conventions to the communicative context

Citation for published version:
Castillo, L, Smith, K & Branigan, H 2019, 'Interaction promotes the adaptation of referential conventions to the communicative context', *Cognitive Science*, vol. 43, no. 8, e12780. https://doi.org/10.1111/cogs.12780

Digital Object Identifier (DOI):
10.1111/cogs.12780

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Cognitive Science

Interaction promotes the adaptation of referential conventions to the communicative context

Lucía Castillo

Department of Psychology, University of Edinburgh

Kenny Smith

Centre for Language Evolution, University of Edinburgh

Holly P. Branigan

Department of Psychology, University of Edinburgh

Author Note

This work was supported by a Becas Chile scholarship to the first author (CONICYT PFCHA/DOCTORADO BECAS CHILE/2013 - 72140201). Address for correspondence: 7 George Square, EH8 9JZ, Edinburgh, United Kingdom. lucia.castillo@ed.ac.uk

Keywords: Convention – Adaptation – Interaction – Alignment – Reference

Abstract

Coordination between speakers in dialogue requires balancing repetition and change, the old and the new. Interlocutors tend to re-use established forms, relying on communicative precedents. Yet linguistic interaction also necessitates adaptation to changing contexts or dynamic tasks, which might favor abandoning existing precedents in favor of better communicative alternatives. We explored this tension using a maze game task in which individual participants and interacting pairs had to describe figures and their positions in one of two possible maze types: a regular maze, in which the grid-like structure of the maze is highlighted, and an irregular maze, in which specific parts of the maze are salient. Participants repeated this task several times. Both individuals and interacting pairs were affected by the different maze layouts, initially using more idiosyncratic description schemes for irregular mazes and more systematic schemes for regular mazes. Interacting pairs, but not individuals, abandoned their unsystematic initial descriptions in favor of a more systematic approach, which was better adapted for repeated interaction. Our results show communicative conventions are initially shaped by context, but interaction opens up the possibility for change if better alternatives are available.

Interaction promotes the adaptation of referential conventions to the communicative context

For people to communicate effectively, they must coordinate their use of language so that speakers express their meanings in ways that their addressees can easily understand. Imagine describing a rendezvous point in a city for an interlocutor: If meeting in central New York, two obvious possibilities suggest themselves – you could exploit the salient grid-structure of Manhattan to describe the meeting place ("the corner of 55th and 7th") or alternatively use salient landmarks as reference points ("two blocks south of Carnegie Hall"). Some of these expressions may be easier than others to produce and understand, as they refer to more salient aspects of the city –for instance, describing a location in Edinburgh's historic Old Town in terms of grid locations or even city blocks is unlikely to be a successful strategy, since the arrangement of the Old Town streets reflects historical happenstance rather than a grid-like organization. How do speakers choose among alternatives? Considerable previous research has emphasized the role of precedents and conventions: Speakers tend to re-use choices –in our example, a grid-based versus landmark-based description strategy– that have previously been successful, and with repeated re-use these choices become established as conventions. But this research does not tell us why a particular choice was made in the first place. More critically, this emphasis on conservatism in dialogue implicitly assumes a conversational context in which relevant features (e.g., the physical context) do not change, and where no new pressures are imposed over the interaction. In the real world, however, conversational contexts and speaker goals change, so that sometimes a previously successful communicative choice may no longer be optimal, undermining rather than supporting coordination. Instead, the need to coordinate may push speakers towards change – and these new choices may themselves then act as precedents in subsequent interaction. On the other hand, a convention can solve a recurrent problem, but not everything in a conversation will be conventionalized. New elements and topics can enter the conversation, and convergence and divergence might coexist, acting over different linguistic elements (Fusaroli et al., 2012). Successful communication therefore requires a balance between repetition and change, the old and the new. But little is known about the mechanisms that determine this delicate balancing act. Moreover, the importance of interaction for this process is also understudied. Faced with the pressure of a dynamic context, will an individual speaker adapt in the same way as a speaker who is interacting directly with a partner? In this study, we explore how the context of communication shapes referential conventions as they initially emerge, and whether and how these initial conventions change through repetition and/or interaction.

1.1 Alignment and the development of conventions

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

Re-using a partner's previous choice makes sense as a strategy for effective communication. The shared experience of the partner's original successful use of that label, scheme, is part of the pair's store of shared knowledge or common ground (Stalnacker, 1978; Clark, 1996). Speakers can appeal to this knowledge when subsequently formulating references for their addressee (i.e., engaging in audience design; Clark & Carlson, 1981), as it represents a safe option (because both speakers agree on its meaning) and it is easily accessible (because it has been recently activated in the conversation; Brennan & Clark, 1996; E. V. Clark, 1997). With repeated re-use, this reference becomes a local convention between the pair, whereby both partners hold a mutual expectation of its subsequent use (Lewis, 1969/2002). This concordant mutual expectation is also strengthened by higher-order expectations between the speakers: Each speaker knows that the other speaker expects them to use this convention. The alignment of linguistic choices between interlocutors may also have a basis in an automatic and resource-free priming mechanism that leads to interlocutors converging on shared alternatives at different linguistic levels (lexical, syntactic, semantic, and situation model; Garrod & Anderson, 1987; Pickering & Garrod, 2004), which in turn entails a common interpretation of the task and facilitates mapping the reused expressions to their referents in the world (Shintel & Keysar, 2009). However, even priming has been found to be influenced by the discourse context and strategic goals of the speakers (Reitter & Moore, 2014; Doyle & Frank, 2016), pointing to an articulation of lower- and higher-order elements in determining the final choices of individuals in interaction.

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

1.2 Determiners of initial conventions

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

In the absence of previous linguistic context, the initial selection of a description scheme might plausibly be shaped by the non-linguistic context. At large scales, this claim has been made for natural languages, where cross-linguistic differences in conventional reference frames have been argued to reflect differences in their functionality in the community's environment (Majid, Bowerman, Kita, Haun, & Levinson, 2004). At the smaller scale of individual interaction, the perceptual context of a conversation involving spatial tasks (such as describing locations in a city or an experimental maze) might play a role in determining which linguistic alternative is preferred. Specifically, salient landmarks or locations can act as either starting points or references to locate other objects or trajectories, constituting a kind of pre-linguistic common ground between interlocutors (Wilkes-Gibbs & Clark, 1992). If our speaker assumes that her interlocutor shares her evaluation of the salience of specific elements in the maze, she might choose to rely on those concrete landmarks to describe her positions (e.g. "the long column on the left", "the one sticking out"). Alternatively, if there are no such landmarks that she can rely on, or if the environment provides another salient conceptualization (e.g., of the maze as a grid), she might opt for a more abstract approach (e.g. "first column, third square down"), depending on her evaluation of the fitness landscape that the perceptual context provides for her descriptions. Importantly, this evaluation would yield similar results whether the speaker was interacting directly with a listener or addressing a future (non-present) interlocutor, as perceptual salience would act similarly on all individuals, with relative independence from their interactional circumstances.

Early in an interaction, we should therefore expect environmental factors (perceptual context, speaker identity, etc.) to outweigh historical factors (current or previous interactions between the same pair) in determining description scheme choices. However, as the interaction progresses, referential expressions that have been successfully used as a coordination tool become salient in the eyes of the interlocutors, shifting speakers' evaluation of the task from a pre-linguistic analysis of the context to an analysis of their common history regarding that context (Garrod, Fay, Lee, Oberlander, & Macleod, 2007; Vogels, Krahmer, & Maes, 2013; Christensen, Fusaroli, & Tylen, 2016). Moreover, as the information content that was initially introduced by one of the speakers becomes shared between the pair, the references that were previously used to refer to this content also become available for both speakers (Xu & Reitter, 2018). Successful references become precedents that speakers can call upon, leading to the development of local conventions. Similarly, an individual speaker lacking concurrent feedback from an audience would become entrained with their own previous expressions, as self-monitoring generates priming effects on the speakers' own production that are equivalent to the effects of priming from an interlocutor's speech (Pickering & Garrod, 2004).

If the communicative situation does not change, the conventions that interlocutors develop should therefore be relatively well adapted to effective communication, because the original description that served as the basis for the convention will generally fit that situation. These initial conventions can then provide a basis for subsequent fine-tuning, which refine or economize initial conventions over the course of repeated use. A well-known example shows how the descriptions of tangram figures, repeatedly exchanged by interlocutors over the course of a game, maintain their core while dropping other attributes, for example going from “a man who looks like he might be pushing something to the right” to “pushing man” (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986). Research in graphical communication has similarly shown that the signs developed in interaction are usually abstracted and refined until an optimal trade-off between ease and economy of production and comprehension has been reached (Fay, Ellison, & Garrod, 2014; Fay, Garrod, & Roberts, 2008; Garrod et al., 2007), showing that even when interlocutors maintain the same conceptual framework, more efficient ways of interaction can be achieved (Garrod & Doherty, 1994; Mills, 2014).

1.3 Alignment and change

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

However, communicative situations frequently change: New referents may come into play, the perceptual context may alter, or the task itself may impose new demands. As the conversational situation changes, established conventions may no longer be appropriate, and successful coordination might require speakers to adapt their referential behavior dynamically to better fit the new situation. Speakers can and quite frequently do change the way that they refer to referents in natural conversation (Healey, 2008). Nevertheless, such changes come at a cost: Even though previous research on conventions suggests that a pairwise convention should not be as strong as a community convention, and so might be defeasible (as the lack of other individuals privy to the convention eliminates an external pressure for conformity; Garrod & Doherty, 1994; Lewis, 1969/2002), all experimental evidence points to a cost associated with abandoning a precedent (Brennan & Clark, 1996; Kronmüller & Barr, 2015; Metzing & Brennan, 2003). In such situations, then, the pressure for change conflicts with a pressure to maintain established precedents.

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

On the other hand, convergence in one aspect of a conversation might not imply across-the-board alignment between speakers. A recent proposal put forward by Xu and Reitter (2018) argues that alignment is bounded by topic shifts, that is, that alignment in one topic or sub-activity in a conversation might be relatively independent from alignment in a different topic or sub-activity, even if the two are intertwined (as frequently occurs in natural conversation). In their proposal, the information content that is initially unbalanced between speakers (as one speaker introduces a new topic) becomes shared as they interact, allowing for an increase in alignment between the speakers as the topic is developed, but not extending outside topic boundaries. Considering a baseline tendency for the alignment and the maintenance of established precedents, we can hypothesize that change and stability might coexist around different topics or sub-activities in an interaction.

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

1.4 The current study

We have argued above that successful collaborative interaction entails a dynamic evaluation that links the speakers' linguistic choices to the demands of the joint action being performed, and to the choices of their co-speaker (Christensen, Fusaroli, & Tylén, 2016; Fusaroli et al., 2012). Importantly, these factors may sometimes exert conflicting pressures: A perceptually salient choice might not be communicatively efficient, and repeating a partner's linguistic choice does not ensure optimal adaptation to the task. Thus, the emergence, establishment, and evolution of linguistic conventions are likely to be the product of multiple competing pressures.

Here we report an experimental study that addressed these issues. Specifically, we sought to test the influence of non-linguistic context on speakers' linguistic choices, and the effect of interaction over these choices. Additionally, we sought to explore the relationship between communicative alignment and adaptation to the task, and the mutual influences between coordination processes acting over different sub-tasks, in this case, different types of referential expressions (namely, spatial description schemes and tangram descriptions). We used a maze game (Garrod & Anderson, 1987; Garrod & Doherty, 1994) to provide a simple and controlled context, in which Individuals and Pairs of participants had to solve a recurrent coordination task. The task required participants to communicate the positions of tangram figures, distributed quasi-randomly in a 'maze' (a series of connected squares), to their partners (co-present in the

Pairs condition, and an imagined future participant in the Individuals condition) and –in the Pairs condition– to locate on their own mazes the positions that were given by their partners (see Figure 1).

------------------------------Insert Figure 1 about here ------------------------------

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

The experimental setup therefore presented participants with a recurrent coordination problem: Even though the mazes were structurally the same from round to round, the tangrams' changing positions meant that participants had to describe different positions on each turn. Hence, although they could re-use tangram descriptions and an overall conceptualization of the maze, they could not re-use specific position descriptions. By presenting participants with this recurrent problem, we aimed to push them towards establishing a consistent description system in order to facilitate processing position information, though participants were not explicitly told to do so. Participants in the Individual condition had to describe tangram positions sequentially for a future participant, and hence received no feedback or input in this process; in contrast, participants in the Pairs condition had to take turns describing and confirming with each other the position of each figure in their respective mazes.

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

With this set-up, we sought to understand, first, what determines people's initial linguistic choices, and what role does perceptual context play in this selection process? If people's initial references are influenced by properties of the physical context, we might expect that Regular mazes, with their visual cues to a consistent underlying structure, would prompt participants to use abstract, systematic descriptions, whereas Irregular mazes, with their visual cues to salient distinct components, would prompt participants to use concrete, figural descriptions (Anderson & Garrod, 1987; Garrod & Anderson, 1987; Garrod & Doherty, 1994).

Second, how does interaction affect participants' adaptation to context? If a repeated need for coordination promotes the use of more efficient descriptions, then we would expect that interacting Pairs of speakers would increasingly come to use descriptions that could be generalized across situations. In other words, they would show an increasing tendency to use

Abstract descriptions, even though this might mean abandoning an established precedent, and they would be faster to complete the task if they did so. This move towards Abstract descriptions should be particularly marked in interacting speakers playing on Irregular mazes, who might initially use Concrete descriptions (influenced by the maze layout) but then – under the conflicting pressure to repeatedly coordinate (see Christensen, Fusaroli, & Tylen, 2016)– should transition to using Abstract descriptions. In contrast, Individual participants working alone, who repeatedly described positions but did not have to coordinate with a partner, should be less likely to change their description strategy. This conservative behavior of Individual speakers would be expected both under a priming-based model of referential choices, as self-monitoring implies speakers prime themselves through monitoring of their own speech, and under an audience design model, as maintaining the same description scheme would be a good strategy to ensure comprehension by a future interlocutor in the absence of in-the-moment feedback (Pickering & Garrod, 2004; Clark & Carlson, 1981; Horton & Gerrig, 2002).

Finally, and more speculatively, how do these processes operate when different types of referring expressions are involved? Specifically, we explore whether alignment in one aspect of a conversation (exchanging descriptions of positions) would be affected by convergence in a different aspect (jointly constructing and reusing a figure's descriptive reference). Following Xu and Reitter (2018), we might assume that convergence and alignment processes are constrained by topic, and thus that alignment onto the same description scheme might be independent from convergence onto a common tangram description. Including repeated reference to tangrams also acted as a check of our interaction manipulation, by allowing us to compare the canonical tangram effect in Pairs versus Individual participants. Following previous research (Branigan, Catchpole, & Pickering, 2011), we anticipated that Pairs of speakers would shorten references to tangrams over consecutive rounds to a greater extent than Individuals.

2. Method

2.1 Participants

128 participants, comprising 64 participants in Pairs[1] (32 pairs; 52 females, 12 males; mean age 21.5 years) and 64 Individual participants (50 females, 14 males; mean age 22.3 years), all University of Edinburgh students, took part in the experiment for payment. All participants were native English speakers. Participants read and signed an informed consent form before taking part in the experiment.

2.2 Materials

We generated four mazes with a maximum size of 7 vertical squares and 7 horizontal squares each (Fig. 2). Two mazes were Regular (high regularity score) and two mazes were Irregular (low regularity score). Maze regularity was measured using the following algorithm: For each square in a maze we calculated the proportion of occupied neighboring squares, considering all 8 (6 for edge squares) surrounding squares (i.e. a square surrounded by occupied squares would obtain a score of 1, a square with only a single neighbor would obtain a score of 0.125), and then took the mean over all squares in the maze to produce a maze regularity score. With this method, mazes that are highly clustered and leave few spaces between squares obtain high regularity scores, while mazes where squares are distributed more loosely throughout the 7x7 space, leaving empty squares in between, obtain lower regularity scores. To ensure our manipulation was linked to regularity differences and not to other properties of the mazes (such as filled vs empty space), Regular and Irregular mazes had the same number of squares (± 1). Controlling the number of squares in the maze in this way ensures that any differences in description schemes between Regular versus Irregular mazes cannot be attributed to difficulty/cognitive load (cf. Anderson & Garrod, 1987, where a "regular" maze comprising a full 6x6 grid totaling 36 squares was compared with an "irregular" maze comprising 24 squares distributed over the same 6x6 grid).

Each Individual or Pair was assigned to one of the four mazes, and used this maze in all three rounds.

-----------------------------Insert Figure 2 about here ------------------------------

We also selected 6 tangrams which were to be the target of descriptions in all conditions. These tangrams were chosen from a set of 30 tangrams that was pre-tested to evaluate the dispersion of their emergent descriptions. Using an online survey, we asked 30 participants (all University of Edinburgh Psychology undergraduates, who took part in exchange for course credit) to describe each figure. Tangrams that elicited the use of the same concept in more than 75% of descriptions were discarded, as this implied they would be too easily identified during interaction in the main experiment. The final set (Fig. 3) was chosen among the figures that were described as "dancing" by at least 20% of participants, which ensured that the tangrams in the set had common features, but that were described using other concepts by at least 50% of participants, allowing them to be adequately differentiated.

-----------------------------Insert Figure 3 about here ------------------------------

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

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

2.3 Design

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

2.4 Procedure

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

Individuals and Pairs of participants played three rounds on the same Irregular or Regular maze. The mazes contained 6 tangram figures in quasi-random positions; in the Pairs condition, the position of the tangrams differed for each partner (see Fig. 1). The same tangram figures appeared in each round, but in different positions. All participants had the same starting point in each round.

In each round, participants worked through the set of 6 tangrams in any order they chose. For each tangram, participants had to describe its position so that a future player (Individuals condition) or their current partner (Pairs condition) could move their icon to that position in the maze (by moving through the paths connecting the squares). In the Individual condition, participants chose a tangram and described its position; they then moved to that position and pressed a key, at which point the selected tangram disappeared from the maze, and they then repeated this procedure until they had cleared all tangrams from the maze. In the Pairs condition, participants selected a tangram and exchanged descriptions of its location in their respective mazes, such that each player could move to the position where the other player's tangram was located. After placing their icons in these complementary positions, both participants pressed a key to ratify their choice, at which point (if the selected positions were correct) the selected tangram disappeared from the maze, and they moved on to describe and select a new tangram. If the selected positions were incorrect, the tangrams did not disappear from the maze; no other feedback was given. A round was finished when the participants managed to make all their tangrams disappear, or when the allocated time of 25 minutes was up (6 Pairs ran out of time in the first round (4 in the Regular condition and 2 in the Irregular condition); no Individuals ran out of time).

2.5 Coding of transcripts

Participants' utterances were initially coded into three categories: tangram description, position reference, or other (greetings, jokes, etc.). The coding of the description schemes used for position references is based on a simplification of the original four description schemes defined by Garrod and colleagues (Line, Coordinate, Figural, and Path: Garrod & Anderson, 1987; Garrod & Doherty, 1994), grouping these schemes into two categories: Abstract and Concrete. Figural (based on salient points or elements, e.g., "the sticking-up bit on the right") and Path descriptions (based on a trajectory between a salient point or element and a goal, e.g., "Two down and one left from the sticking-up bit") were classified as Concrete ('Type 1 sub-language' in Healey, 2008), as they require each position to be described in relation to a specific element in the maze, and are therefore dependent on the actual configuration of the maze. Line (based on vertical or horizontal elements in the maze, e.g., "The second horizontal row, all the way to the right") and Coordinate descriptions (established as the intersection of a vertical and a horizontal element, e.g., "2nd column, 4th row") were classified as Abstract ('Type 2 sub-language' in Healey, 2008), as they rely on an abstract grid-like pattern which is independent from the actual configuration of the maze, and could therefore be applied to any maze without change. All position descriptions were coded, totaling 1634 descriptions from 32 Pairs, and 1035 descriptions from 64 Individual participants (some Pairs produced more than one description per position, usually after the first description was not uniquely identified).

Each position description was coded for alignment, i.e. whether it used the same description scheme as the preceding description produced by the other participant in the Pair, or by the same participant in the Individual condition. In order to capture change of linguistic choices over time as well as conceptual change, our alignment code considered the finer-grained set of contrasts originally proposed by Garrod and colleagues (i.e. a Figural description followed by another Figural description would be counted as alignment, but a Figural description followed by a Path description would not). For Pairs, if the same speaker made more than one consecutive position reference, the comparison was nonetheless performed against the partner's previous reference (i.e., we did not consider self-alignment in Pairs). For Individual participants, alignment was measured as self-alignment with their own previous description.

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

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

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

Box 1. Example of co-constructed tangram description by two participants, A and B. All words in bold are counted as one description.

3. Results

We present our results[2] in four sub-sections. We first consider how our main experimental manipulation (Regular vs Irregular maze layouts) influenced participants’ use of Abstract vs. Concrete descriptions. We then investigate the relationship between use of Abstract vs. Concrete descriptions and time taken to complete the task. We next examine how participants’ use of descriptions was influenced by their partner’s previous linguistic behavior (i.e., alignment). Finally, we examine how participants’ tangram descriptions were affected by interaction and repetition.

3.1 Choice of description schemes

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

We used a logistic multilevel model (lme4 package in R, Bates, Maechler, Bolker, & Walker, 2015) to evaluate the effect of our layout manipulation on the use of Abstract or Concrete description schemes. We included Interaction condition (Pairs vs Individuals), Layout regularity and Round number as fixed effects, plus their interactions. The random structure was the maximal structure justified by our design (following Barr, Levy, Scheepers, & Tily, 2013), including random intercepts for Pair/Individual and for Participant nested in Pair[3] (to account for the initial variation between pairs and between subjects), and random slopes for Pair/Individual over Round number (to account for the differential effect of round number over the different Pairs or Individual participants). P-values were obtained through likelihood ratio tests for each parameter against a model without that parameter, using the function *mixed* in the Afex package (Singmann, Bolker, Westfall & Aust, 2017). Predicted probabilities were generated using the function *predict* in the lme4 package.

-----------------------------Insert Figure 4 about here ------------------------------

There was a significant main effect of Interaction condition (Pairs vs Individuals) on the use of Abstract vs. Concrete descriptions: Participants in Pairs were more likely to use Abstract description schemes than Individual participants (predicted probabilities 68.7% vs 29.5%, in round 1). Critically, there was a significant effect of Layout: Participants tended to use significantly more Abstract descriptions in Regular layouts than Irregular layouts (predicted probabilities 71.3% vs 33.4%, in round 1). Round number also had a significant effect on the use of Abstract descriptions, with participants using more Abstract descriptions as they played more rounds (predicted probabilities 56.1% for round 1, 62.9% for round 2, and 67.0% for round 3). Moreover, we found a significant interaction between Interaction condition (Pairs vs Individuals) and Round number: Participants in the Pairs condition were significantly more likely to use Abstract descriptions as they played more rounds (predicted probabilities rising from 66.4% in round 1 to 80.5% in round 3, across Layouts), compared to the Individual condition (predicted probabilities rising from 41.1% to 48.3%, across Layouts). No other interactions were significant (see Table 1).

-----------------------------Insert Table 1 about here ------------------------------

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

We ran a generalized linear model to predict the overall time taken to complete the task (i.e. all 3 rounds) from the Layout participants were assigned to, their Interaction condition, and their Use of abstract descriptions (the mean proportion of Abstract descriptions produced by each Pair as a numeric value[4]), plus their interaction. To account for the impact of the different task demands over the two Interaction conditions, time was scaled by Interaction condition[5].

The model showed significant main effects of Interaction condition and Use of abstract descriptions: After scaling, Individual participants were faster than Pairs in completing the task; and crucially, the Use of abstract descriptions was associated to faster times across Interaction conditions and Layouts. Moreover, there was a significant interaction between Interaction condition and Layout: Pairs were slower than Individuals in Regular layouts (see Table 2).

To test the possibility that time taken to complete the task was affected by alignment, we ran a further model that predicted total time per Pair or Individual (scaled by Condition, 3 rounds) from Use of abstract descriptions, Alignment rate per Pair/Individual, Interaction condition, and their interaction. There was no main effect of Alignment rate on the time taken to complete the task, $F(1,55)= 1.9982$, $p=0.736$.

------------------------------Insert Table 2 about here -------------------------------

------------------------------Insert Figure 5 about here -------------------------------

3.3 Alignment on description schemes

We ran a binary logistic mixed-effects model predicting alignment on position description schemes (i.e., whether participants used the same description scheme as used in their partner's previous description, in the Pairs condition; or their own previous description, in the Individual condition) based on Interaction condition, Layout, Round number, and their interaction; we used the same coding scheme as reported in section 3.1, however the random intercept for participant nested in Pair was removed as its variance and standard deviation were zero.

We found a main effect of Interaction condition: Individual participants were more highly aligned than participants in Pairs (though recall that Individual participants were self-aligning, whereas Pairs were aligning with their interlocutor) (predicted probabilities 89.4% vs 76.4%, in round 1); and a main effect of Round number: Alignment scores increased significantly with each round played (predicted probabilities 81.3% on round 1, 87.6% on round 2, 89.3% on round 3). No interactions were significant (see Table 3).

An additional analysis of the data for each Condition separately revealed a differential response by Pairs and Individuals. Whereas Individuals showed only a significant main effect of Round number ($X^2=7.25$, $p=0.007$), Pairs showed both a main effect of Round ($X^2=9.55$, $p=0.002$) and an interaction between Layout and Round ($X^2=4.04$, $p=0.04$), with Pairs in Regular layouts showing a greater increase in alignment as they played more rounds, compared to Pairs in Irregular layouts. Although this effect was not strong enough to appear in the full model as a 3-way interaction, it reflects the difference in the alignment trajectory of Pairs according to their initial conventions: Pairs in Regular layouts maintained their initial conventions, consistently increasing their alignment levels as they played more rounds, but Pairs in Irregular layouts abandoned their initial descriptions, temporarily misaligning as a consequence (see Fig. 6, Irregular layout).

------------------------------Insert Figure 6 about here -------------------------------

------------------------------Insert Table 3 about here -------------------------------

3.4 Tangram description length

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

We ran a linear mixed-effects model to predict the (log-transformed) number of words used per tangram description per Pair[6] or Individual participant, with Interaction condition, Round number, Layout regularity, Tangram order (1 to 5[7]) and their interactions as fixed effect predictors. We included a random intercept for Pair/Individual and a random slope for Pair/Individual over Round number (the inclusion of a random slope for tangram order resulted in failure to converge). P-values were obtained from the *mixed* function in the afex package.

We found a significant main effect of Round number ($F(1,243)=144.62$, $p<0.001$), with participants using fewer words to describe tangrams with each new round played; and a significant main effect of Tangram order ($F(1,781)=6.10$, $p=0.01$), with participants using fewer words to describe each additional tangram within each round. There were no main effects of Interaction condition or Layout. Importantly, we found a significant interaction between Interaction condition and Round number ($F(1,243)=168.82$, $p<0.001$): With each new round played, Pairs' descriptions shortened while Individuals' did not. The interaction between Interaction condition and Tangram order was also significant ($F(1,781)=25.35$, $p<0.001$) : Descriptions of each new tangram in round 1 were shorter in the Pairs condition, but not in the Individual condition. Notice however that there was also a 3-way interaction between Interaction condition, Round number, and Tangram order ($F(1,778)=31.82$, $p<0.001$): Pairs' descriptions were shorter for each new tangram on round 1, but this effect diminished in later rounds (see Table 4).

There was also a significant interaction between Round number and Tangram order ($F(1,778)=13.55$, $p<0.001$): Tangram order had a larger effect at round 1 than at later rounds, since descriptions shortened during round 1 but were relatively flat within subsequent rounds; and between Interaction condition and Layout ($F(1,93)=6.01$, $p=0.02$): Pairs in Regular layouts produced longer descriptions than Pairs in Irregular layouts in round 1, first tangram, compared to Individuals. As this effect disappeared after a few descriptions, it suggests the additional linguistic coordination needed to establish an abstract description system initially affected participants' verbosity[8].

------------------------------Insert Figure 7 about here -------------------------------

------------------------------Insert Table 4 about here -------------------------------

In a final exploratory analysis, we carried out correlation tests between the random effect estimates from the tangram description length model and the alignment model, to investigate the possibility of a relationship between the referential contraction process and participants' alignment with their partner's previous description scheme use (Pairs condition) or their own previous use (Individual condition). If participants who produce more aligned descriptions are also producing references that are more contracted, and inversely participants who produced fewer aligned descriptions are producing less contracted references, this would suggest a relationship exists between alignment in one type of descriptions and the conventionalization process implied in the referential reduction of a different type of descriptions. The analysis was performed on the random effects estimates, rather than directly on the variables, in order to obtain a better assessment of the independent variability (i.e. not attributable to Interaction condition or Layout) of the two variables, while reducing the uncertainty of the measures. The correlation analysis revealed no significant relationship between residual mean length of tangram descriptions and residual alignment scores (correlation of random intercept estimates (measuring initial alignment and tangram description length scores), $r=-0.048$, $p=0.711$; correlation of random slope estimates for Round number (measuring progress of alignment and tangram description lengths over rounds), $r=-0.138$, $p=0.289$; correlation of random intercept estimates from alignment model and random slope estimates for tangram length model (measuring the relation between initial alignment scores and progress of tangram description length over rounds), $r=0.062$, $p=0.636$).

4. Discussion

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

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

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

The regularity of maze layouts did not affect the overall time taken to complete the task. However, participants who used more Abstract descriptions were significantly faster to complete the task than participants who used more Concrete descriptions, and Individuals were faster than

Pairs (after scaling of time by Interaction condition). We now consider our results and their theoretical implications in turn.

4.1 Referential contraction and the role of interaction

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

Importantly, Pairs' tendency to contract their tangram references was not affected by alignment or misalignment in their position descriptions. A Pair could be transiently misaligned, i.e., in the process of abandoning a concrete description scheme and moving towards an abstract one, while at the same time maintaining their coordinated referential contraction process by converging on the same tangram descriptions (reflected in the absence of a correlation between alignment rate and tangram description length). This pattern reflects an interesting dual processing of different semantic stimuli, where speakers can process, contract, and align on one set of referential expressions (in this case, the descriptions of the tangrams themselves) while at the same time misalign and diverge on a different set of expressions (in this case, the descriptions of the tangrams' positions), pointing towards a structuring of alignment processes around topics or sub-tasks (Xu & Reitter, 2018).

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

Individuals, lacking confirmation of the success of their past descriptions, seemed to put more effort (in terms of number of words) into subsequent descriptions.

4.2 The influence of context in the emergence of conventional references

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

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

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

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

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

4.3 Change from concrete to abstract description schemes: who, how and why?

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

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

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

This type of shift cannot be explained by the effect of salience or alignment alone, and instead points towards an interaction between dialogical alignment and a dynamic adaptation to a changing context of use (Christensen, Fusaroli, & Tylén, 2016). In our experiment, the appropriateness or "fit" of a description scheme with respect to a specific maze layout generated an initial contextual pressure, but the repetition of the task made some schemas more efficient than others over time, creating an additional and competing pressure. Pairs (but not Individuals) adapted to this new pressure by migrating towards a type of schema that could be more easily applied to new mazes or new positions in the same maze, namely Abstract schemas. In this sense, the language used in the Pairs adapted to its circumstances of use (Lupyan & Dale, 2016), where a system that facilitated repeated interaction was needed.

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

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

The fact that Individual speakers showed no change throughout the game in terms of their position descriptions is compatible with more than one account. An alignment-based explanation would predict self-monitoring to act as a source of alignment within the speaker's own language processing system, as their linguistic output is examined for consistency between the intended output and the form actually produced. Individual speakers' consistency could also be explained in terms of cognitive economy, hence egocentric processing, as the introduction of a different perspective when this was not required by the task would certainly imply more processing effort. An audience design explanation (i.e., based on considerations of a partner's mental states) would similarly predict consistency, as the speakers' model of a future interlocutor cannot be contrasted with an actual co-present interlocutor, and is therefore fixed as an idealized model in their evaluation of the task.

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

4.4 From pairs to communities

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

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

The link between pairwise interactions and the community level of language change is directly addressed by Garrod and Doherty (1994), who argued (in line with Lewis, 1969/2002) that speakers turn to the schema they have most frequently encountered in the past when they face a new interaction. In this sense, the influence of their past interactions would be additive, with previously encountered variants competing with respect to the proportions of use in which they have been experienced. However, the description schemes emerging from the communities in Garrod and Doherty (1994) are not conventional in Lewis' terms, as there is no need for beliefs about the behavior of other members of the community. As speakers do not need to be aware of the existence of a community at all to align on the most frequently encountered variant (though beliefs can modulate this tendency; see Branigan et al., 2011; Fehér, Wonnacott, & Smith, 2016), Garrod and Doherty posit a more mechanistic explanation, where the global level emerges out of the sum of pairwise interactions.

Our results suggest, however, that alignment and frequency of use do not act alone in determining which variants will be adopted by a speaker, but that their content or 'value in context' also plays a role. Tamariz et al. (Tamariz, Ellison, Barr, & Fay, 2014), using data from an artificial language experiment, explained the process by which these conventions form as a consequence of the interplay between two biases: the tendency to re-use an existing variant, and a preference for better (i.e. simpler or more communicatively functional) variants. The interplay between the two biases leads individuals to retain their own previously used variant unless they encounter a superior variant in terms of content[9], in which case this new variant would be adopted. Under this account, communicative interaction, and in particular the evaluation of the communicative utility of alternative communicative tokens, is crucial to the spread of new variants and the formation of population-level conventions. Similar processes can be identified in our experiment. Participants in Pairs tended to align on the same description scheme variant, determined in the first place by the individual evaluation of the descriptions' properties against the context. However, as the interaction progressed, participants identified a superior variant in terms of its appropriateness in context and ease of use in repeated exchanges, and this variant was then adopted and stabilized in subsequent games.

5. Conclusions

Alignment has been considered to be the main driver in the establishment of conventional references. However, as our results show, participants need to take into account more than just their partner's contributions when interacting in a given context. Both Individuals and Pairs of speakers produced references that were adapted to their context of use, but only interacting Pairs were able to dynamically switch from an established reference to a new alternative when the task demanded a change. Communicative alignment in context needs to consider both linguistic interaction and content functionality in order to reach an equilibrium that is not simple repetition, but informed convergence.

Acknowledgements

We thank Gregory J. Mills and Alisdair Tullo for programming support.

References

Anderson, A., & Garrod, S. C. (1987). The dynamics of referential meaning in spontaneous conversation: Some preliminary studies. In R. G. Reilly (Ed.) *Communication failure in dialogue and discourse* (pp. 161–183). Amsterdam, The Netherlands: Elsevier.

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, 46, 391–418. https://doi.org/10.1006/jmla.2001.2815

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3). https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01

Branigan, H. P., Catchpole, C. M., & Pickering, M. J. (2011). What makes dialogues easy to understand? *Language and Cognitive Processes*, 26(10), 1667–1686. https://doi.org/10.1080/01690965.2010.524765

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 22(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Christensen, P., Fusaroli, R., & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, 146, 67-80. https://doi.org/10.1016/j.cognition.2015.09.004

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

Clark, H. H. & Carlson, T. B. (1981) Context for comprehension. In J. Long, A. Baddeley (Eds.) *Attention and Performance, IX* (pp. 313-331). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, l-39. https://doi.org/10.1016/0010-0277(86)90010-7

Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. *Cognition*, 64, 1–37. https://doi.org/10.1016/S0010-0277(97)00010-3

Doyle, G., & Frank, M. C. (2016). Investigating the sources of linguistic alignment in conversation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vol. 1, pp. 526-536).

Fay, N., Ellison, T. M., & Garrod, S. (2014). Iconicity: From sign to system in human communication and language. *Pragmatics & Cognition*, 22(2), 244–263. https://doi.org/10.1075/pc.22.2.05fay

Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B-Biological Sciences: Biological Sciences*, 363(1509), 3553–3561. https://doi.org/10.1098/rstb.2008.0130

Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language,* 91, 158-180. https://doi.org/10.1016/j.jml.2016.06.002

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: quantifying the benefits of linguistic coordination. *Psychological science*, 23(8), 931-939. https://doi.org/10.1177/0956797612436816

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2), 181–218. https://doi.org/10.1016/0010-0277(87)90018-7

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition* 53, 181-215. https://doi.org/10.1016/0010-0277(94)90048-5

Garrod, S., Fay, N., Lee, J., Oberlander, J., & Macleod, T. (2007). Foundations of representation: where might graphical symbol systems come from? *Cognitive Science*, 31(6), 961–987. https://doi.org/10.1080/03640210701703659

Garrod, S., & Pickering, M. J. (2009). Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science*, 1(2), 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Grice, H.P. (1975). Logic and Conversation. In Cole et al. *Syntax and semantics 3: Speech arts* (pp. 41-58). Retrieved from http://www.ucl.ac.uk/ls/studypacks/Grice-Logic.pdf

Healey, P. G. T. (1997). Expertise or Expert-ese?: The Emergence of Task-Oriented Sub-Languages. *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, 301–306.

Healey, P. G. T. (2008). Interactive Misalignment: The Role of Repair in the Development of Group Sub-languages. *Language in Flux*, 13–39.

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. https://doi.org/10.1016/S0749-596X(02)00019-0

Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness monitoring and ambiguity. *Psychological Science*, 13(3), 207–212. https://doi.org/10.1111/1467-9280.00439

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1(1-12), 113-114. https://doi.org/10.3758/BF03342817

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4(3), 343–346. https://doi.org/10.1037/h0023705

Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language*, 83, 1–19. https://doi.org/10.1016/j.jml.2015.03.008

Lewis, D. K. (1969/2002). *Convention : a philosophical study*. Oxford, UK: Blackwell Publishing.

Lupyan, G., & Dale, R. (2016). Why Are There Different Languages? The Role of Adaptation in Linguistic Diversity. *Trends in Cognitive Sciences*, 20(9), 649–660. https://doi.org/10.1016/j.tics.2016.07.005

Majid, A., Bowerman, M., Kita, S., Haun, D. B., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, 8(3), 108-114. https://doi.org/10.1016/j.tics.2004.01.003

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49(2), 201–213. https://doi.org/10.1016/S0749-596X(03)00028-7

Mills, G. J. (2014). Dialogue in joint activity: Complementarity, convergence and conventionalization. *New Ideas in Psychology*, 32(1), 158–173. https://doi.org/10.1016/j.newideapsych.2013.03.006

Misyak, J., Noguchi, T., & Chater, N. (2016). Instantaneous Conventions: The Emergence of Flexible Communicative Signals. *Psychological Science*, 956797616661199. https://doi.org/10.1177/0956797616661199

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, 27(2), 169-190-226. https://doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, 36(4), 329–47. https://doi.org/10.1017/S0140525X12001495

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, 76, 29-46. https://doi.org/10.1016/j.jml.2014.05.008

Shintel, H., & Keysar, B. (2009). Less Is More: A Minimalist Account of Joint Action in Communication. *Topics in Cognitive Science*, 1(2), 260–273. https://doi.org/10.1111/j.1756-8765.2009.01018.x

Shirouzu, H., Miyake, N., & Masukawa, H. (2002). Cognitively active externalization for situated reflection. *Cognitive Science*, 26(4), 469–501. https://doi.org/10.1207/s15516709cog2604_3

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *afex: Analysis of Factorial Experiments*. R package version 0.16-1. https://CRAN.R-project.org/package=afex

Stalnaker, R. C. (1978). *Assertion*. Oxford, UK: Blackwell Publishers Ltd.

Tamariz, M., Ellison, T. M., Barr, D. J., & Fay, N. (2014). Cultural selection drives the evolution of human communication systems. *Proc. R. Soc. B*, 281. https://doi.org/10.1098/rspb.2014.0488

Tarenskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: salience, absoluteness, and consistency. *Frontiers in Psychology*, 6, 1703. https://doi.org/10.3389/fpsyg.2015.01703

Van Der Wege, M. M. (2009). Lexical entrainment and lexical differentiation in reference phrase choice. *Journal of Memory and Language*, 60(4), 448–463. https://doi.org/10.1016/J.JML.2008.12.003

Vogels, J., Krahmer, E., & Maes, A. (2013). Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production. *Language and Cognitive Processes*, 28. https://doi.org/10.1080/01690965.2012.682072

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31(2), 183–194. https://doi.org/10.1016/0749-596X(92)90010-U

Xu, Y., & Reitter, D. (2018). Information density converges in dialogue: Towards an information-theoretic model. *Cognition*, 170, 147-163. https://doi.org/10.1016/j.cognition.2017.09.018

Tables

Table 1

Results of model of description scheme use

Predictor	Log-odds Betas	SE	X^2 and p values
Intercept	-0.2657	0.6894	-
Interaction condition	4.7584	1.6146	12.44, p=0.0004***
Layout	4.4427	1.5325	10.80, p=0.0010**
Round number	0.9044	0.2588	7.29, p=0.007**
Interaction condition * Layout	0.4531	2.6952	0.03, p=0.87
Interaction condition * Round number	1.9711	0.8416	9.27, p=0.002**
Layout * Round number	0.7870	0.7235	1.75, p=0.19
Interaction condition * Layout * Round number	0.5229	1.0358	0.25, p=0.62

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

Table 2

Results of model of time taken to complete task

Predictor	Betas	SE	F and p values
Intercept	-0.03213	0.04336	-
Layout	0.02385	0.08671	(1,81) 1.8086, p=0.178
Interaction condition	0.22885	0.08671	(1,82) 6.0091, p=0.014*
Use of abstract descriptions	-0.36886	0.10800	(1,83) 9.0564, p=0.002 ***
Layout * Use of abstract descriptions	-0.33609	0.21601	(1,79) 0.2365, p=0.626
Layout * Interaction condition	0.38394	0.17342	(1,78) 5.1824, p=0.022*
Interaction condition * Use of abstract descriptions	-0.34908	0.21601	(1,80) 0.2873, p=0.592
Layout * Interaction condition * Use of abstract descriptions	-0.38760	0.43202	(1,77) 1.0092, p=0.315

Note: Betas and Standard Error (from generalized linear model *glm* in the *lme4* package), and Chi-squared and P-values (from likelihood-ratio tests using *lrtest* function). Time estimates are shown in minutes. Model was fitted using a Gamma distribution (link=log), with deviation-coded variables Layout (Irregular-Regular) and Interaction condition (Individuals-Pairs). Use of Abstract Descriptions was mean-centered. Raw time was scaled by Interaction condition before entering the model.

Table 3

Results of model of aligned descriptions

Predictor	**Log-odds Betas**	**SE**	**X^2 and p values**
Intercept	2.3330	0.2590	-
Interaction condition	-1.1741	0.4567	7.04, p=0.008**
Layout	0.2932	0.4434	0.44, p=0.51
Round number	1.2937	0.3525	22.60, p<0.001***
Interaction condition * Layout	0.4411	0.8872	0.25, p=0.62
Interaction condition * Round number	-0.2937	0.4196	0.53, p=0.47
Layout * Round number	0.5096	0.3983	1.71, p=0.19
Interaction condition * Layout * Round number	1.3125	0.7862	2.74, p=0.10

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

Table 4

Results of model of tangram description length

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

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

Figures

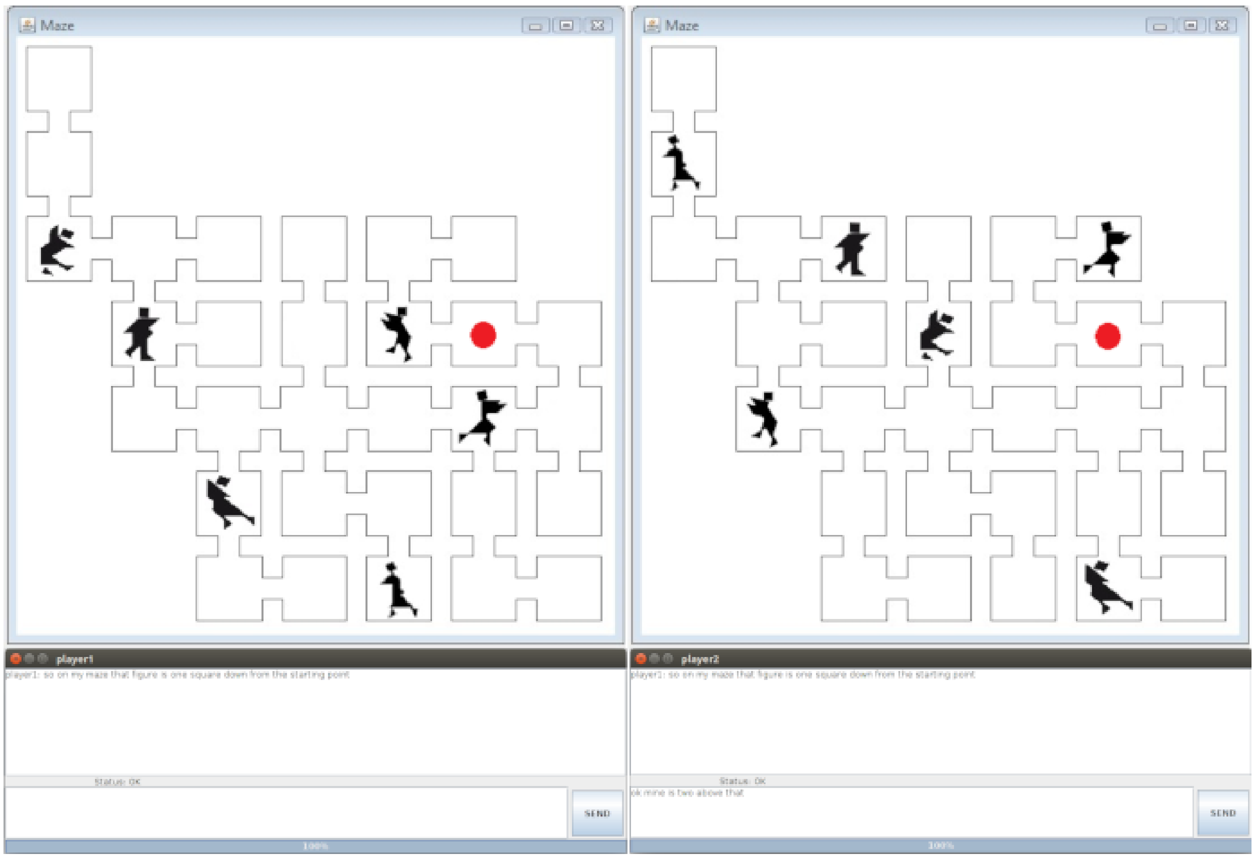

Figure 1. Mazes for Player A (upper left) and B (upper right) in a round of the game, as it appeared to players – players also saw a window featuring the chat tool interface which they could use to communicate with their partner (lower left and right). The player's own position in the maze is indicated by the dot, and there are 6 tangram figures randomly positioned within the maze – note that the two players have the same tangrams, but in different positions. This maze has a high regularity score of 0.821; see Figure 2 for irregular mazes.

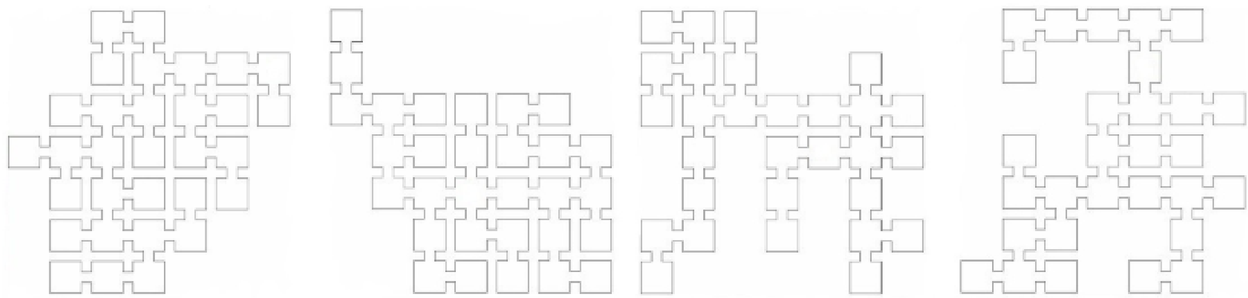

Figure 2. Regular (left side) and Irregular (right side) mazes. Regular mazes have regularity scores of 0.864 and 0.821, while Irregular mazes scores are 0.672 and 0.668 (e.g. a regularity score of 0.864 indicates that on average, 86.4% of the neighbors of a square in that maze are occupied, i.e. roughly 7 out of 8). As the figure shows, Irregular mazes contain more spaces between occupied squares, which leads to multiple salient configurational features (protrusions, indentations and clusters).

Figure 3. Tangram set.

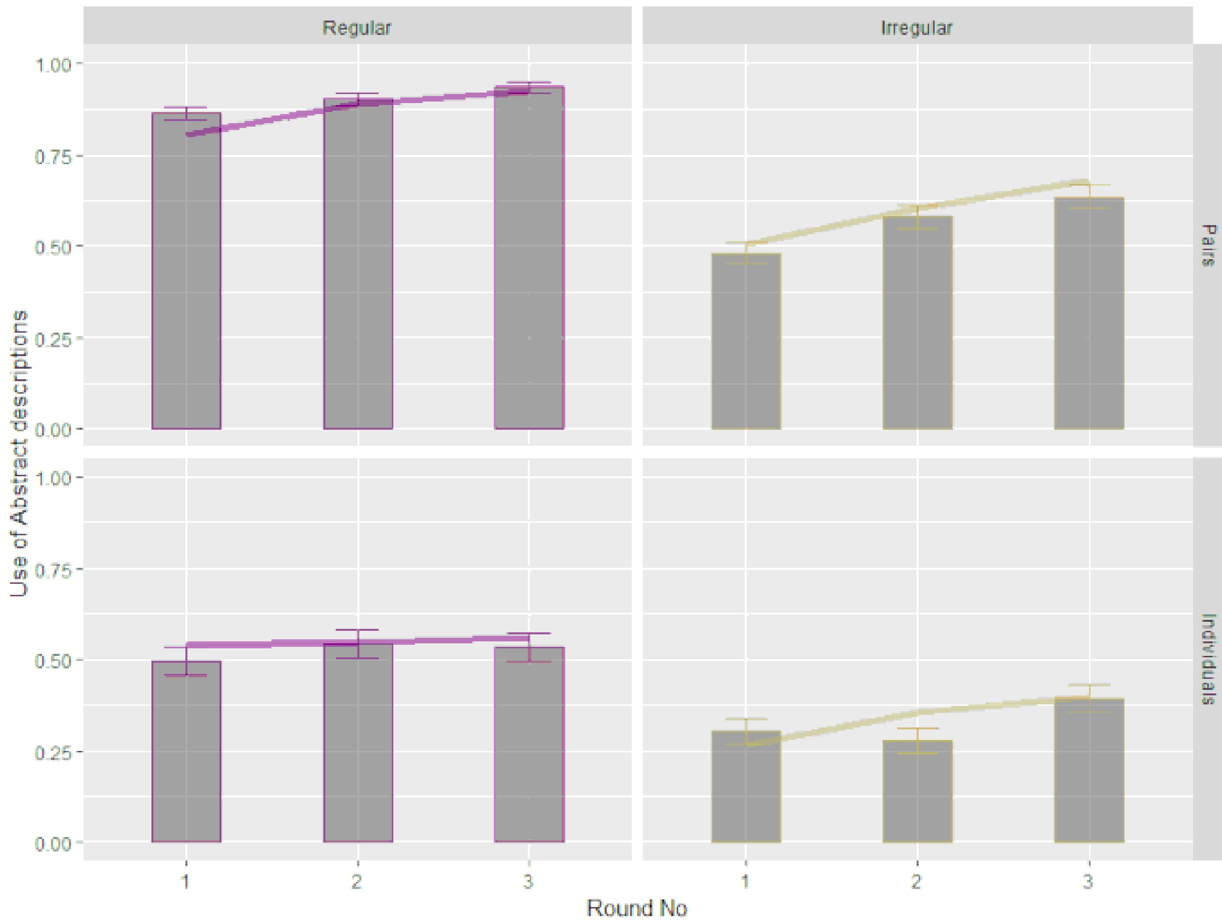

Figure 4. Proportion of abstract descriptions over the three rounds of interaction, in Pairs (upper panels) and Individuals (lower panels), for Regular (left panels) and Irregular mazes (right panels). Colored lines show model predictions, bars show raw data, error bars show SE.

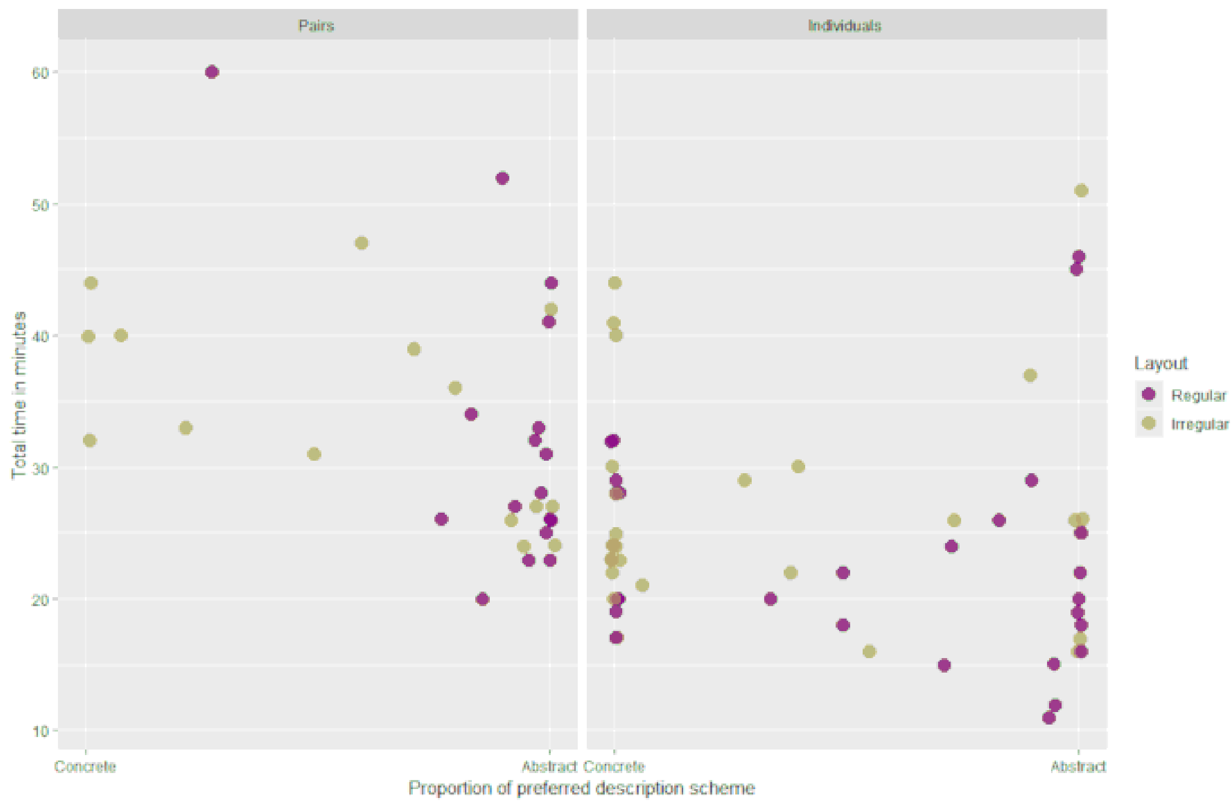

Figure 5. Time taken to complete the experiment (i.e., 3 rounds) for Pairs (left panel) and Individuals (right panel), as a function of the proportion of Abstract and Concrete descriptions they produced (x axis goes from absolute use of Concrete descriptions on the left, to absolute use of Abstract descriptions on the right, for each panel). Each point represents an Individual or Pair, and points are colored according to the maze Layout they played on (Regular: dark shaded, Irregular: light shaded).

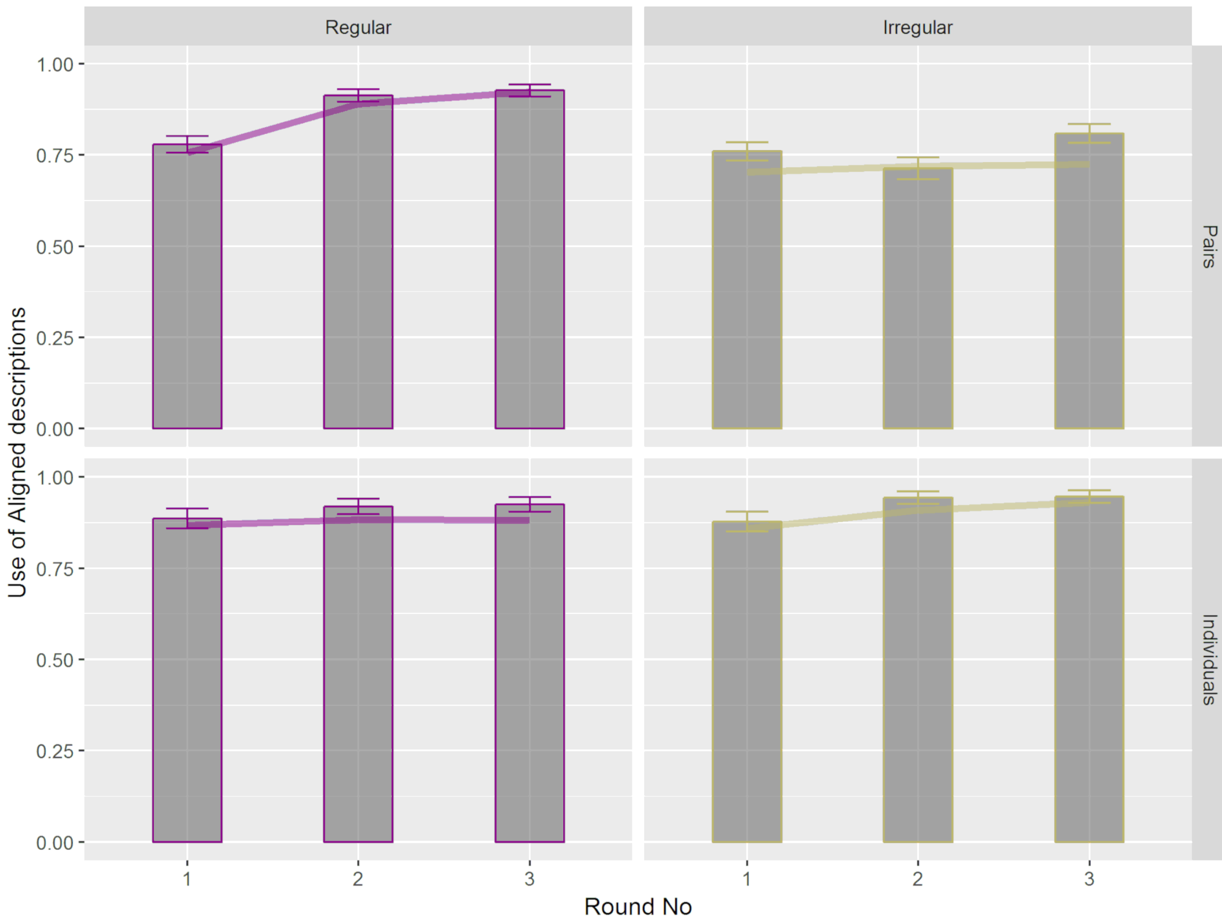

Figure 6. Proportion of aligned descriptions over three rounds of interaction, in the Pairs (upper panels) and Individuals (lower panels) conditions, for Regular mazes (left panels) and Irregular mazes (right panels). Colored lines show model predictions, bars show raw data, error bars show SE.

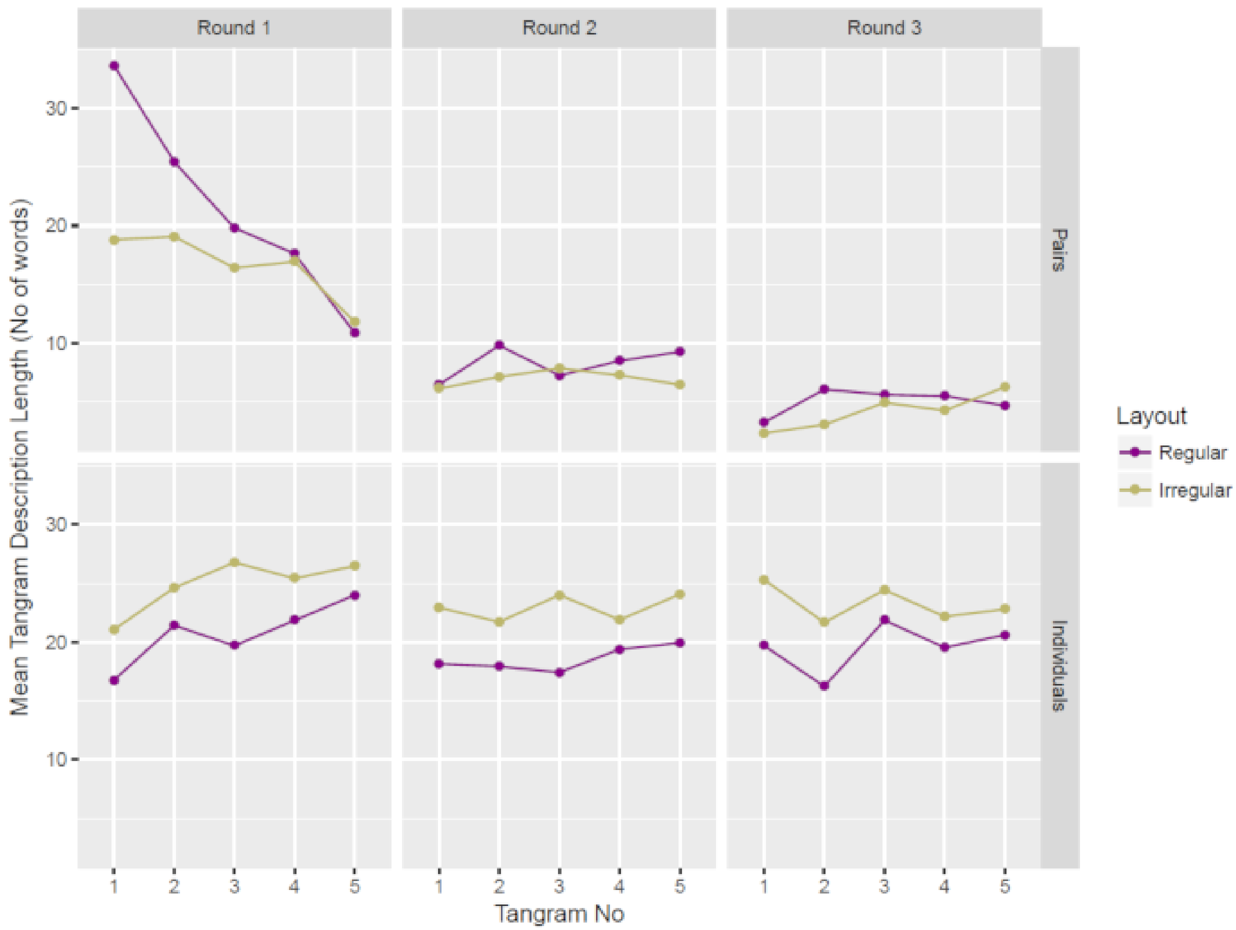

Figure 7. Mean number of words per tangram description per round, in Pairs (upper panels) and Individuals (lower panels), for Regular (dark shaded) and Irregular (light shaded) Layouts.

Footnotes

[1] Data analysis considered the identification of each individual speaker in a Pair; for the purpose of statistical analysis, individual participants were nested in their respective pair. Our design employed the same number of participants on each interaction condition, considering this factor and the relevance individual experience can have on this type of task.

[2] Data and analysis scripts are available at https://osf.io/gzxma/

[3] Individual participants were assigned a unique "Pair" number, to allow for the fit of the same model over Pairs and Individual participants.

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

[5] Scaling procedure involves dividing all values in each Interaction condition by the root mean square of that condition.

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

[7] Participants decided arbitrarily the order of the tangrams they described; there was no pre-established order. Tangram order corresponds to the arbitrary order used by each Pair in a given round.

[8] An additional analysis on position description length revealed an overall decrease in turn length (number of words) with Round, across Conditions and Layouts ($F(1,199)=15.25$, $p<0.001$), with no other significant predictors. However, as the positions described changed in each round, and considering we cannot determine whether specific positions were repeated between rounds due to the quasi-randomization process, these data can give only a rough estimate of the phenomenon, and therefore

---

cannot be directly compared with the referential contraction process that occurs over specific descriptions for particular tangrams.

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