Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Structural priming in artificial languages and the regularisation of unpredictable variation

Olga Fehér [a,*], Elizabeth Wonnacott [b], Kenny Smith [a]

[a] School of Philosophy, Psychology and Language Sciences, University of Edinburgh, United Kingdom
[b] Division of Psychology and Language Sciences, University College London, United Kingdom

ARTICLE INFO

Article history:
Received 16 September 2015
revision received 30 May 2016
Available online xxxx

Keywords:
Structural priming
Artificial language learning
Unpredictable variation

ABSTRACT

We present a novel experimental technique using artificial language learning to investigate the relationship between structural priming during communicative interaction, and linguistic regularity. We use unpredictable variation as a test-case, because it is a well-established paradigm to study learners' biases during acquisition, transmission and interaction. We trained participants on artificial languages exhibiting unpredictable variation in word order, and subsequently had them communicate using these artificial languages. We found evidence for structural priming in two different grammatical constructions and across human-human and human-computer interaction. Priming occurred regardless of behavioral convergence: communication led to shared word order use only in human-human interaction, but priming was observed in all conditions. Furthermore, interaction resulted in the reduction of unpredictable variation in all conditions, suggesting a role for communicative interaction in eliminating unpredictable variation. Regularisation was strongest in human-human interaction and in a condition where participants believed they were interacting with a human but were in fact interacting with a computer. We suggest that participants recognize the counter-functional nature of unpredictable variation and thus act to eliminate this variability during communication. Furthermore, reciprocal priming occurring in human-human interaction drove some pairs of participants to converge on maximally regular, highly predictable linguistic systems. Our method offers potential benefits to both the artificial language learning and the structural priming fields, and provides a useful tool to investigate communicative processes that lead to language change and ultimately language design.

Introduction

All human languages exhibit a shared set of organizing principles or structural properties, sometimes known as design features (Hockett, 1960). Recently there has been a surge of experimental studies exploring how these properties of language can be explained in terms of individual-level processes of language learning and language use, and how biases in these processes shape language change and ultimately, language design (see Kirby, Griffiths, & Smith, 2014, for review). These studies focus on characteristics of language that are shared across all languages (e.g. the use of combinatorial and compositional processes to construct complex signals: Kirby, Cornish, & Smith, 2008; Kirby, Tamariz, Cornish, & Smith, 2015; Verhoef, 2012) and explore the extent to which these reflect the biases of individual learners and the processes which come into play when languages are transmitted between individuals via learning. Here we consider

[*] Corresponding author.
E-mail addresses: olga.feher@ed.ac.uk (O. Fehér), kenny.smith@ed.ac.uk (K. Smith).

http://dx.doi.org/10.1016/j.jml.2016.06.002
0749-596X/© 2016 The Author(s). Published by Elsevier Inc.

another mechanism by which language is shaped: communicative interaction. Language is learnt and transmitted in a rich communicative context, therefore people's moment-to-moment linguistic choices during interaction will ultimately shape language structure over longer timescales. We use a novel combination of experimental techniques to look at how some of these processes, namely priming and convergence during interaction, influence language use and how this affects the structure of language. We take unpredictable variation as our test case because of its long history of experimental investigation via artificial language paradigms (discussed in detail below). We combine techniques and insights from this literature with methods from structural priming to investigate the relationship between priming, communicative interaction and regularisation. This may inform us about the role of communicative interaction in driving language change and ultimately shaping language design.

Regularisation of unpredictable variation

Natural languages are inherently variable at all levels of linguistic structure: phonetic, morphological, syntactic, semantic and lexical variation are all common. Linguistic variation tends to be predictable, that is, conditioned either on grammatical or social context (Givón, 1985). Such conditioning may be deterministic, as for the English plural marker allophones [s] (as in "cats"), [z] (as in "dogs") and [ɪz] (as in "horses") whose occurrence is conditioned on the preceding sound. Other variation is probabilistic in nature: in some situations speakers are more likely to produce certain variants than in others. Most sociolinguistic variation is of this type. For instance, the pronunciation of -ing English (as in "finding", "running") typically takes one of two forms: [ɪŋ]("-ing") or [ɪn]("-in"), and speakers' choice varies according to the formality of the situation and the speaker's gender (Fischer, 1958) as well as their social status (Shuy, Wolfram, & Riley, 1967). Even in these probabilistic cases variation is therefore conditioned and somewhat predictable, although the conditioning factors may be complex and subtle. Truly unpredictable, unconditioned 'free' variation seems to be relatively rare in language. This fact about language stands in need of explanation, particularly since one could imagine that tracking complex conditioning rules would place more demands on memory and online processing during speech than randomly sampling from the set of available variants.

One explanation for the scarcity of unpredictable variation in natural language is that it is eliminated during language acquisition due to strong learner biases against such variation. This has been studied in the laboratory using artificial language paradigms. In these paradigms, participants are exposed to a miniature, experimenter-designed language and then tested on their sensitivity to violations of the regularities embodied in the language using a range of judgment, comprehension and production measures (see Gomez & Gerken, 2000, for a general review of artificial language learning methods). Artificial language paradigms are a well-established means to obtain maximum experimental control over learners' input (Aslin, Saffran, & Newport, 1998), and several studies have provided evidence that artificial languages are processed similarly to natural languages (e.g. Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Wonnacott, Newport, & Tanenhaus, 2008). Moreover, artificial language paradigms have a long history as a tool for exploring statistical or distributional learning and learners' biases in language acquisition. These paradigms have been used extensively to study word segmentation (e.g. Saffran, Aslin, & Newport, 1996), word learning (e.g. Smith & Yu, 2008; Yu & Smith, 2007), the learning of grammatical categories (Frigo & McDonald, 1998; Gerken, Wilson, & Lewis, 2005), and the acquisition of phonology (Chambers, Onishi, & Fisher, 2010) and syntax (Reeder, Newport, & Aslin, 2013; Wonnacott et al., 2008; Wonnacott, Boyd, Thomson, & Goldberg, 2012) in both adults and children.

A common finding of these studies is that human language learners are able to track the statistics of their input and use this information to infer underlying linguistic structure. However, as mentioned above, artificial language learning paradigms can also be used to explore how learning biases shape language, by seeing how learners respond to linguistic features that are uncharacteristic of natural languages, such as unpredictable variation. Various experiments have explored how learners deal with an artificial language with synonymous forms whose use varies unpredictably (unlike in a natural language). These studies focus on whether learners will mirror this unpredictable usage when producing the language themselves, or whether they will eliminate the variation, either by dropping competing forms or by conditioning variant use on context. Pioneering artificial language learning experiments demonstrated that children eliminate unpredictable variation during learning (Hudson Kam & Newport, 2005, 2009), suggesting that the absence of unpredictable variation in human languages may be a direct consequence of biases in child language acquisition. In contrast, adult learners are more likely to reproduce the probabilistic usage of variants and probability match the statistics of their input (Hudson Kam, 2009; Hudson Kam & Newport, 2005, 2009). However, input complexity seems to play a role: adults tend to regularise more when the input is both unpredictable and complex (e.g. when there are multiple unpredictably varying synonymous forms) but can acquire quite complex systems of conditioned variation (e.g. where there are multiple synonymous forms whose use is lexically or syntactically conditioned: Hudson Kam, 2015; Hudson Kam & Newport, 2009). There is suggestive evidence that conditioning facilitates the learning of variability by children, although they are less adept at acquiring conditioned variation than adults (Hudson Kam, 2015; Samara, Smith, Brown, & Wonnacott, submitted for publication), and children's preferences for regularity are reduced if the learning task is simplified, e.g. by mixing novel function words and grammatical structures with familiar English vocabulary (Wonnacott, 2011).

As well as being restructured by the biases of individual language learners, languages may also be shaped by processes of transmission: where languages are repeatedly passed between human learners, weak biases in learning can be amplified. Artificial language learning has also been used to study this process. Smith and Wonnacott (2010) demonstrated that when an artificial language exhibiting unpredictable variation is transmitted across a chain of five adult learners (using an iterated learning paradigm where the language produced by one learner after training becomes the target language for the next learner in a chain of transmission), unpredictability is entirely eliminated due to the cumulative effects of weak individual-level biases in favor of regularisation. This finding is in line with a growing body of experimental work showing that iterated learning may serve to restructure languages (see Kirby et al., 2014, for review).

Transmission between individuals provides one mechanism by which individual-level processes (i.e. language learning) can shape the properties of natural languages. However, language is not only transmitted from one user to the next, but continuously used for communication. The process of interaction may also exert pressures which shape language: interaction may have lasting effects on the user's language system. In addition, the observation of such communicative interactions forms the input to child language acquisition. The capacity for moment-to-moment pressures operating during language use to shape linguistic systems is well-established in linguistic theory: frequency and distinctiveness shape the use of linguistic forms, their mental representation, and therefore ultimately the structure of language (e.g. Bybee, 2001, 2006; Garrod & Pickering, 2013; Wedel, 2007). Moreover, the ways in which speakers bend linguistic conventions to meet ever-changing communicative needs drives the continual production of linguistic innovations and constitutes the engine for language change (e.g. Croft, 2000; Heine, 1997).

What happens when languages exhibiting unpredictable variation are used during communication? Again, artificial language learning paradigms provide a useful tool to explore this process. Smith, Fehér, and Ritt (2014) taught participants artificial languages that exhibited random variation (presence or absence of a grammatical marker), and then had them communicate dyadically using this language. The members of each communicating pair were trained on languages which used the grammatical marker in different proportions, yet despite the mismatch between their linguistic systems and the unpredictability of their linguistic input, participants rapidly converged during interaction, coming to use the grammatical marker with the same frequency. Furthermore, participants tended to converge on regular (i.e. non-varying) or predictably variable systems (Fehér, Ritt, & Smith, in preparation; Smith et al., 2014. This suggests that interaction may act to reduce linguistic variation, and contribute to the constrained patterns of variation we see in all natural languages. However the mechanisms by which this process occurs are unclear. In the current paper, we explore the role of syntactic priming in this process.

Priming and convergence during interaction

During linguistic interaction, interlocutors modify their behavior to match that of their partners via a process known as accommodation or alignment (Coupland, 2010; Pickering & Garrod, 2004). According to Pickering & Garrod (2004), linguistic convergence occurs because of a simple priming mechanism: hearers' linguistic representations are activated during comprehension, which increases the likelihood of them using the same forms or structures when they speak. Priming occurs at various levels of linguistic representation: phonetic (Giles, Coupland, & Coupland, 1991), lexical (Brennan & Clark, 1996; Garrod & Anderson, 1987), semantic (Garrod & Anderson, 1987; Garrod & Clark, 1993), and structural (Bock, 1986; Gries, 2005). Structural priming, where interlocutors match the syntactic structure of their partners' utterances (see Pickering & Ferreira, 2008, for review) is particularly interesting for several reasons. Firstly, from the point of view of communicative interaction, sentence structure is somewhat independent of the propositional content of utterances – languages typically provide a number of structurally distinct means of conveying a given idea. Furthermore, structural priming can be used to probe the mental representations underpinning language. For instance, the observation that structural priming occurs across modalities (from comprehension to production: e.g. Branigan, Pickering, & Cleland, 2000; Pickering, Branigan, Cleland, & Stewart, 2000) strongly suggests that these processes access shared mental representations of syntactic structure. Critically, structural priming has been shown to occur even when prime and target sentences share no lexical items, although it may be increased by shared lexical items (the so-called 'lexical boost'), which is taken as evidence of abstract representations at the level of syntactic categories. For this reason, there has been particular interest in demonstrating structural priming in children, particularly in light of the theories predicting that early syntactic representations might not be fully abstract (Tomasello, 2000). It is now well established that children do show structural priming, responding to primes with no shared lexical items, from at least around 3 years of age (Huttenlocher, Vasilyeva, & Shimpi, 2004; Peter, Chang, Pine, Blything, & Rowland, 2015; Rowland, Chang, Ambridge, Pine, & Lieven, 2012; Savage, Lieven, Theakston, & Tomasello, 2003). In fact, children may not show the same lexical boost as adults, which has recently been interpreted as showing that this phenomena stems from explicit awareness of the repetition of lexical items across prime and target sentences, which is present only in adults (Peter et al., 2015).

In terms of its role in natural dialog, structural priming has been found to enhance communicative success (Pickering & Garrod, 2006). It has been demonstrated in natural dialog (Levelt & Kelter, 1982; Schenkein, 1980; Weiner & Labov, 1983), in corpora (Gries, 2005) and it has been shown experimentally in numerous communicative situations (e.g. Branigan et al., 2000; Branigan, Pickering, McLean, & Cleland, 2007). Even in non-interactive situations, people are primed by their own linguistic productions: when asked to describe a scene, they are more likely to reuse a syntactic form they just used in an unrelated sentence they read aloud (Bock, 1986) or heard and repeated (Bock & Griffin, 2000). Moreover, in sentence-completion tasks, people have been shown to align on their previous spoken (Bock & Griffin, 2000;

Branigan et al., 2000; Hartsuiker & Westenberg, 2000) or written productions (Pickering & Branigan, 1998).

To study structural alignment in dialog while maintaining control over the language participants were exposed to, Branigan and colleagues introduced the confederate-scripting technique (Branigan et al., 2000). This involves confederate and a naive participant taking turns describing scenes using one of two structures (either a Prepositional Object construction, as in "gave the cake to the burglar", or a Double Object construction, as in "gave the burglar a cake"). The confederate's choice of syntactic form was pre-determined, and the key measure of interest was the extent to which the participant changed their use of the two structural alternatives given the forms produced by the confederate. Branigan and colleagues found that participants were primed by the usage of the confederate's previous syntactic form.

In addition to the largely automatic priming that has been demonstrated by psycholinguistic research which focuses on the cognitive mechanisms underlying language production and comprehension, socially mediated alignment has also been a focus of a number of studies in sociolinguistics. An early finding was that people adjust their linguistic behavior to their interlocutors more when they perceive their partner more favorably (Giles, Taylor, & Bourhis, 1973; Giles & Powesland, 1975). More recently, Balcetis & Dale (2005) used the confederate scripting technique to show greater syntactic alignment with a likeable confederate than with an unlikeable confederate. Anti-alignment can also be used as a mark of disaffiliation with a speaker (Bourhis, Giles, Leyens, & Tajfel, 1979; Doise, Sinclair, & Bourhis, 1976). The degree of intentionality behind this type of behavioral alignment or anti-alignment is under debate; a recent study found syntactic alignment to be largely automatic and independent of social perceptions but the *degree* of alignment to be socially mediated (Weatherholtz, Campbell-Kibler, & Jaeger, 2014). Behavioral mimicry can, in turn, influence our perception of our interaction with others (Chartrand & Bargh, 1999). We will return to these issues in the discussion.

The present study: structural priming in unpredictably variable artificial languages

In this paper we present a novel experimental technique combining methods for studying structural priming (scripted and non-scripted interaction) with artificial language learning. Our first goal was to investigate whether structural priming would be apparent, and whether, as in natural languages, it would occur even across utterances with no shared lexical items (although we also look for a potential lexical boost, as reported in the literature). Our second goal was to explore the relationship between these priming processes and the elimination of unpredictable variation from the language system. Smith et al. (2014) and Fehér et al. (in preparation) found clear evidence that communication between participants led them to converge on a more regular language. One possibility is that this occurred via the process of reciprocal structural priming, i.e. whereby the two participants repeatedly reinforced their usage of a particular variant and thereby converged on a more regular system. There may also be other more general aspects of interaction which create or augment a bias for regularisation. Language is a system of communicative conventions: part of its communicative utility comes from the fact that interlocutors tacitly agree on what words and constructions mean, and deviations from the 'usual' way of conveying a particular idea or concept are therefore taken to signal a difference in meaning (e.g. Clark, 1988; Horn, 1984). This suggests that producing unpredictable linguistic variation during communication might be counter-functional – the interlocutors of a speaker who produces unpredictable variation might erroneously infer that the alternation between several forms is intended to signal *something* (i.e. is somehow conditioned on meaning). Language users might implicitly or explicitly know this, and reduce the variability of their output during communicative interaction. If this is the case, we should expect to see some regularisation even when there is no opportunity for reciprocal priming, and even when any structural priming that does occur should actually act against regularisation. There is experimental evidence suggesting that this strategic reduction of variability might occur during communicative language use. Perfors (2016) trained adult participants on a miniature language exhibiting unpredictable variation in the form of (meaningless) affixes attached to object labels. While in a standard learn-and-recall condition participants reproduced this variability quite accurately, in a modified task where they were instructed to attempt to produce the same labels as another participant undergoing the same experiment at the same time (who they were unable to interact with), they produced more regular output (producing the most common affix on approximately 80% of labels, rather than on 60% as seen during training). This could be due to reduced pressure to reproduce the training language 'correctly' due to the changed focus on matching another participant, or it could reflect reasoning about the rational strategy to use in this semi-communicative scenario. The two experiments reported in this paper are designed to allow us to tease apart the contribution of these two mechanisms – reciprocal priming and strategic reduction in variability – to regularisation during interaction.

We adopt a naming-and-matching paradigm similar to that used by Branigan et al. (2000), and combine it with methods borrowed from the unpredictable variation literature, training participants on artificial languages that exhibit unpredictable variation in word order. We present two experiments looking at two different artificial languages exhibiting different types of syntactic variation and exploring different types of communicative context. In Experiment 1, participants attempt to learn artificial languages where noun phrases consist of a noun and either a numeral or an adjective, based on a paradigm that has been used to show that learners' biases for certain word orders mirror biases in the distribution of these word orders across the world's languages (Culbertson, Smolensky, & Legendre, 2012; Culbertson & Newport, 2015). In Experiment 2, we look at priming of a more complex transitive construction (from Wonnacott et al., 2008), and also manipulate participants' communication partners (participants interact with another participant or the computer) and their beliefs about their partner (whether or not participants interacting with the computer believe they are in fact interacting with another participant). These manipulations allow us to explore whether regularisation is affected by (1) the opportunity for reciprocal priming (i.e. in human-human interaction), which could potentially lead to convergence on a more regular language and/or (2) the participant's belief about their partner, i.e. whether they produce a more regular language when they believe they are interacting with another human, reflecting the strategic avoidance of unpredictable variation when communicating with another person.

Combining experimental paradigms from two distinct fields (unpredictable variation and structural priming) offers substantial potential benefits to both: artificial language learning methods offer psycholinguists complete control over participants' linguistic experience prior to interaction, and can in principle be used to probe representations of any structural feature of interest, including those absent from participants' natural language (or indeed from any natural language). In return, methods from priming provide the artificial language learning community with a powerful tool to study the representations that participants form during learning novel linguistic input. Our focus in this paper is on using this method to explore how utterance-by-utterance processes of priming, alignment and convergence[1] in language use might shape the structure of linguistic systems, taking unpredictable linguistic variation as a test-case for exploring the relationship between language learning, language use, and language design.

Experiment 1

We borrow an artificial language paradigm used by Culbertson and colleagues with adult and child learners (Culbertson et al., 2012; Culbertson & Newport, 2015). The target language consists of four sentence types: Numeral-Noun, Noun-Numeral, Adjective-Noun and Noun-Adjective. Note that there are two types of noun phrase (involving two modifier types, numerals or adjectives) which both exhibit unpredictable variability in word order. Culbertson and colleagues explored whether participants would regularise within and across modifier types given such a language, and how this was affected by the probabilities of the different word orders in their input. Their central finding was that participants tended to regularise more when the statistically dominant word orders were harmonic (i.e. exhibiting the same ordering of modifier and noun for both modifier types, i.e. Numeral-Noun and Adjective-Noun, or Noun-Numeral and Noun-Adjective), which mirrors the observation that harmonic orders are more frequent across the world's languages (e.g. Culbertson et al., 2012, p. 309 report that, in a sample of 851 languages, 670 have harmonic ordering of numerals and adjectives with respect to the noun). Here, we borrow an input condition where adult participants did not exhibit strong regularisation biases, specifically one where Numeral-Noun and Noun-Adjective were the dominant orders, on the basis that this offered us the best chance of seeing flexible word order use by our participants, and therefore the potential for priming. The choice of an artificial language featuring two related phrase types also allows us to study structural priming both within and across grammatical categories.

We modify the artificial language methodology used by Culbertson and colleagues by adding an interactive stage to their learning paradigm. As in a standard artificial language study, learners are first exposed to the language and are then tested on their ability to immediately recall that language (a stage of the experiment we refer to as *recall test 1*). They then play a director-matcher naming game with the computer using the artificial language (the *interaction* stage). Finally, participants are asked to attempt to recall the training language a second time (*recall test 2*), allowing us to observe any lasting effect of interaction on their memory of the artificial language. In order to compress the entire learning and interaction procedure into a single session, we used the smaller, more transparent set of lexical items from Culbertson & Newport (2015), in order to minimize the amount of time spent on training vocabulary. Across all three stages (recall 1, interaction, recall 2) we are interested in participants' use of the possible word orders. In particular, we looked for (1) regularisation (during recall 1, interaction, or recall 2), which would be evidenced by a reduction in the variability of the artificial language (as indexed by, e.g., over-production of the majority word order, or a reduction in the total entropy of the language), and (2) structural priming during interaction, which would be evidenced by a tendency to reuse the word order used by their interlocutor (Glermi, their alien language tutor) in the immediately preceding trial.

$$
\begin{aligned}
S_{Adj} &\rightarrow Adj\ N \quad (p = 0.3)\\
S_{Adj} &\rightarrow N\ Adj \quad (p = 0.7)\\
S_{Num} &\rightarrow Num\ N \quad (p = 0.7)\\
S_{Num} &\rightarrow N\ Num \quad (p = 0.3)
\end{aligned}
$$

$$
\begin{aligned}
N &\rightarrow \{\text{grifta, mauga, nerka, slerga}\}\\
Adj &\rightarrow \{\text{bluth, flurf, sprat}\}\\
Num &\rightarrow \{\text{dof, threz, fortch}\}
\end{aligned}
$$

Fig. 1. The grammar of the target language in Experiment 1. Each sentence consists of a noun and a modifier, with adjectives being mainly (but not always) postnominal, and numerals being mainly (but not always) prenominal (as indicated by sentence probabilities, *p*).

[1] The term "alignment" has been used both to describe convergence on shared representations, perspectives or behaviors (Pickering & Garrod, 2004), and also in reference to matching a partner's previous linguistic production (Branigan, Pickering, Pearson, McLean, & Brown, 2011). In this paper, we use the terms "convergence" and "priming" to differentiate between the two levels of analysis. By convergence we mean that interacting participants produce languages with similar statistical properties, and by priming we mean the matching of moment-to-moment linguistic behavior between communication partners.

Please cite this article in press as: Fehér, O., et al. Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language* (2016), http://dx.doi.org/10.1016/j.jml.2016.06.002

Method

Participants

Twenty participants (5 males, 15 females, mean age 22.8) were recruited from the University of Edinburgh's Student and Graduate Employment service to take part in a miniature language communication experiment. Participants were paid £7 for their participation.

Procedure

Participants were seated in isolation in sound-attenuated booths, and worked through a computer program (written in Python using Psychopy: Peirce, 2007) which presented and tested them on an artificial language (named Verblog, see Fig. 1), and then allowed them to use that language to communicate with the computer.[2] The language was text-based: participants observed pictures of novel objects that differed either in a surface feature (such as texture or color) or in number (2, 3 or 4 objects of the same type) together with text descriptions, and entered their responses by keyboard. Each stage of the experiment was introduced by an alien character, Glermi, who was their language tutor and communicative partner.

Participants progressed through an eight-stage training and testing regime. All stages except for interaction (stage 7) are based closely on the paradigm from Culbertson & Newport (2015), originally designed to be run over two days with child learners.

Stage 1) Noun training: Participants viewed pictures of four novel objects (in greyscale, of roughly equal size) along with nonsense nouns (*grifta*, *mauga*, *nerka* and *slerga*). Each presentation lasted 3.5 s: for the first 0.5 s the picture of the object was shown alone, then the noun was presented above the object for 3 s. Participants received 20 such training trials, in random order, with each noun being presented 5 times.

Stage 2) Noun comprehension: Participants were prompted with labels and asked to identify the correct object. On each trial, participants were presented with a single noun (one of the four listed above) and an array featuring all four novel objects (position within the array randomised), and were asked to select (by keypress) the correct object given the label, with no time limit on their response. Participants could track their level of success with a counter at the side of the screen: correct selections were followed by a rising, cheerful tone and the addition of 10 points to their running total; incorrect responses were followed by a falling, dull tone and no points. Participants received 20 such training trials, in random order, with each noun being presented 5 times.

Stage 3) Noun testing: Participants were presented with a picture of an object, without accompanying text, and were asked to provide the appropriate label (20 trials total, each object presented five times in random order).

Stage 4) Sentence training: Participants were exposed to pictures of objects paired with two-word descriptions of those objects. The same four objects used in previous stages were used here. However, the objects were either presented singly and differing in color, texture or patterning (introducing the adjectives: each object could be blue, fluffy or spotty), or in groupings of multiple objects in their greyscale form (introducing the numerals: each scene featured two, three or four instances of an object). The accompanying two-word description featured an adjective and noun or a numeral and noun, as appropriate. Following Culbertson & Newport (2015), the adjectives and numerals were selected so as to be relatively transparently related to their meaning (*bluth* for blue, *flurf* for fluffy, *sprat* for spotty; *dof* for two, *threz* for three, *fortch* for four). The ordering of modifier and noun was randomised independently for each trial: adjectives preceded nouns with probability 0.3 (and otherwise followed them, yielding a tendency for N-Adj order); numerals preceded nouns with probability 0.7 (and otherwise followed them, yielding a tendency for Num-N order) – note that since word order was generated on the fly at each trial, the ratio of the various word orders differed slightly between participants, but in all cases followed the intended statistical pattern. Each presentation lasted 4.5 s: for the first 0.5 s the picture of the object(s) was shown alone, then the appropriate two-word description was presented above the object(s) for 4 s. Participants received 48 such training trials, in random order, with each combination of noun and modifier (4 nouns × 6 modifiers) being presented twice (since word order is generated independently for each presentation, the word order for the two presentations of a given combination of noun and modifier may have differed).

Stage 5) Sentence comprehension: Participants were prompted with two-word descriptions and asked to select the appropriate object(s), with no time limit being imposed on responses. Each trial consisted of a two-word description, featuring a noun and either an adjective or a numeral, together with an array of four objects. On adjective trials, the test array featured all four combinations of two objects (the target object plus a foil object) and two adjectives (the target adjective plus a foil adjective), the foil object and adjective being randomly selected; similarly, on numeral trials the test array featured all four combinations of two objects and two numerals. The word order used in the presented description was randomised for each trial (according to the same probability distribution used in sentence training, and with word order at every trial being generated independently, as in sentence training). As during noun comprehension, participants could track their level of success: scores were reset to zero at the start of sentence comprehension, correct selections were followed by a rising, cheerful tone and the addition of 10 points to their running total, with a falling, dull tone and no points for incorrect responses. Participants received 48 such comprehension trials, in random order, each combination of noun and modifier presented twice.

Stage 6) Individual recall test 1: Participants viewed the same 24 images (4 objects × 6 modifiers, order randomised, each image presented twice for a total of 48 test trials) without accompanying text and were asked to enter the appropriate description. Unlike during interaction (see below), no feedback was provided on the descriptions participants provided.

[2] We are grateful to Jennifer Culbertson for providing original code and stimuli on which this experiment was based.

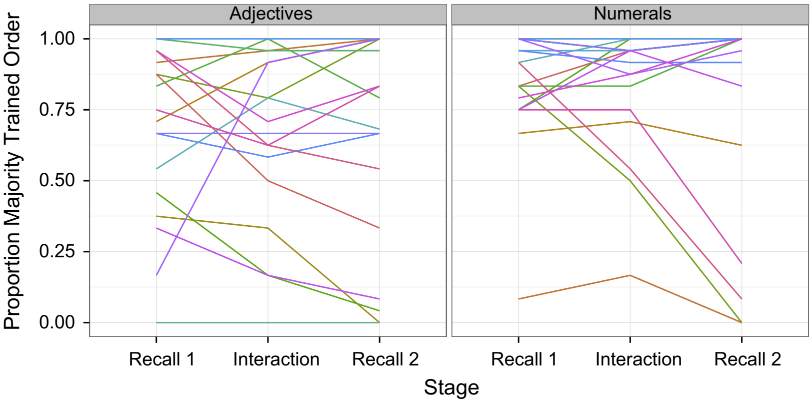

Fig. 2. Proportion of sentences produced in the majority trained order (N-Adj or Num-N) during the initial recall test, interaction, and the final recall test, for both adjectives and numerals. Each line represents a participant, participants are represented by lines of the same color across the two trial types. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Stage 7) Interactive testing: This stage exploits the comprehension and recall trial types described above, to allow the participant to interact with Glermi. Participants played a director-matcher game in which they alternated describing objects for Glermi, and selecting objects based on Glermi's descriptions. When acting as matcher, the participant received a two-word description from Glermi and was required to select the appropriate picture from an array of four objects; all details of matcher trials were identical to sentence comprehension trials (i.e. word order was randomly generated at each trial using the same underlying probability distribution as in training, the composition of the matcher array followed the same constraints, participants received feedback and tracked their updated score). When directing, participants were presented with a picture (drawn from the set of 24 possible combinations of noun and modifier) and prompted to type the description so that Glermi could identify it. This description was then evaluated by the experimental software, playing the part of Glermi: the computer considered the 8 possible legal descriptions for the 4 objects in its matcher array (two word orders for each scene, the computer's matcher array being generated in exactly the same way as a genuine participant's matcher array) and simply selected the image whose description was closest (according to Levenshtein distance) to the description provided by the director; in the case of two or more descriptions being equally close, the computer matcher selected randomly among those candidates. Participants then received feedback in the same way as in matcher trials: a rising tone and 10 points for successful trials, a falling tone and 0 points for unsuccessful trials. Participants played 96 such communication games, such that they directed and matched twice for each possible image (split into two blocks of 48 trials, order randomised within a block, the participant being equally likely to direct or match on the first trial, and the roles alternating for the remainder of the block).

Stage 8) Individual recall test 2: Participants viewed the same 24 images as in recall test 1 (order randomised, each object presented twice for a total of 48 trials) without accompanying text and were asked to enter the appropriate description "to check you remember all the correct descriptions"; as in recall test 1, no feedback was provided on the descriptions participants provided.

Analysis

All analyses were carried out using logit mixed-effects regressions with the maximal random effects structure (by-participant random intercepts and random slopes for all within-subjects manipulations[3]). For analyses dealing with word order, we automatically identified the word order of participants' judgements, by identifying the noun in each production and whether it was sentence-initial (N-Mod order) or sentence-final (Mod-N order).

Results

Communicative success

Performance during the communicative portion of the task was extremely high throughout and varied relatively little across conditions (mean in first half of interaction = 98.3% of trials correct, *SE* = 0.4%; mean in second half of interaction = 98.4%, *SE* = 0.4%).

Word order use

Figs. 2 and 3 show the full data for word order use across recall test 1, interaction, and recall test 2: Fig. 2 shows the timecourse of use across the various phases for the two modifier types separately, and Fig. 3 shows each participant's behavior for both modifier types at each stage. The majority of our participants produced variable linguistic output during the individual recall test immediately following training, although a number of participants were fully regular (i.e. used a single word order for each modifier type), particularly for Numeral trials. As can be seen in Fig. 3, most of the individuals who regularised moved to a more regular version of the trained language

[3] In a small number of cases, these models failed to converge, in which case we simplified the random effect structure: typically removing interactions between random slopes was sufficient. In all such cases the non-converging models showed the same pattern of effects as the reduced models.

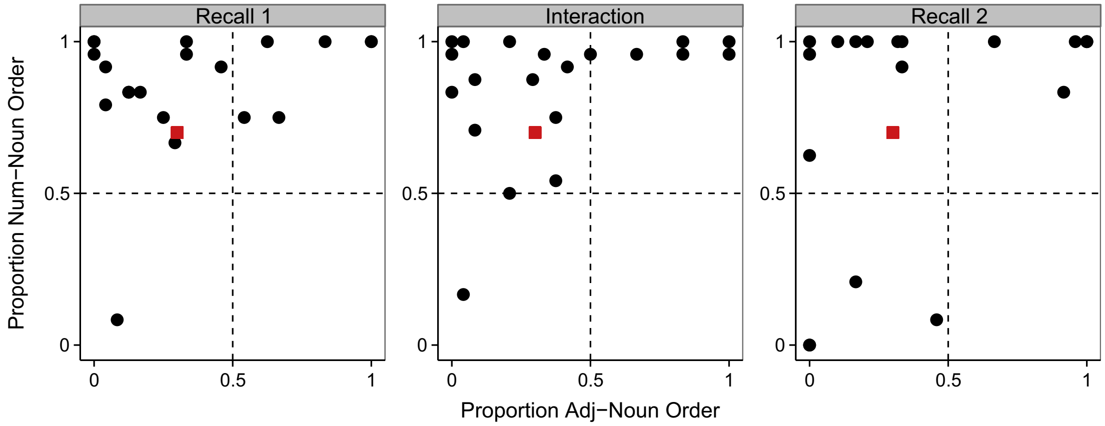

Fig. 3. Proportion of sentences produced in the Noun-Final order for Adjective and Numeral trials by each individual participant in the three experimental stages. Each dot represents a participant, and the mean training proportion is given by the red square. The dashed lines indicate 50% (maximally variable) word order usage. English word order would be represented by dots in the upper right corner. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

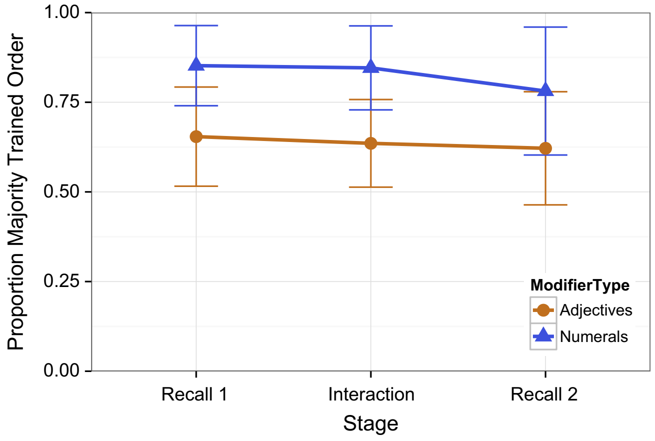

Fig. 4. Mean proportion of sentences produced in the majority trained order (N-Adj or Num-N) during the initial recall test, interaction, and the final recall test, for both adjectives and numerals. Error bars indicate 95% CIs.

(marked by a red square), and did not simply produce English-typical word order (upper right quadrant).

A logit regression exploring the effect of Modifier (dummy coded, taking Adjective trials as our reference level) and test block (recall test 1, interaction, recall test 2, with recall test 1 as intercept), with each trial coded as being in the majority trained order or not (i.e. N-Adj for adjective trials; Num-N for numeral trials), indicates that participants were on average producing adjectives in the trained proportion (as indicated by the model intercept: $\beta = 0.990$, $SE = 0.651$; this is not significantly different from observed usage during training, where the log odds of N-Adj order are 0.847, non-significant difference at $p = .826$). However, participants were marginally more likely to use the majority trained order on numeral trials ($\beta = 1.720$, $SE = 0.937$, $p = .066$), and produced Num-N order significantly more often than it was encountered during training (log-odds of producing Num-N = 2.717, $SE = 0.500$, significantly higher than the trained odds ratio of 0.847, $p < .001$). Visual inspection of the individual data in Fig. 3 suggests that this tendency towards use of the majority trained word order increased during interaction and the second recall test, although inspection of the fitted model (or the aggregated data show in Fig. 4[4]) indicates this effect was not reliable (as indicated by n.s. effects of experiment stage and n.s. interactions between stage and modifier type: largest $\beta = 0.506$ seen in the interaction between numeral trials and interaction stage, $SE = 0.318$, p = .110) – however, the entropy analysis below speaks more directly to this question.

Fig. 5 plots the entropy of participants' word order choices for the two modifier types and also for the lan-

[4] In this figure and throughout the paper, we use code from http://www.cookbook-r.com/Graphs/Plotting_means_and_error_bars_(ggplot2) to provide within-subject error bars, using the technique from Cousineau (2005) with the correction from Morey (2008).

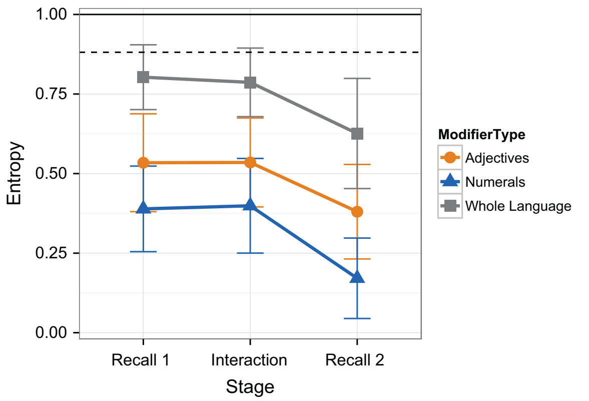

Fig. 5. Entropy of word order, with higher entropy indicating more variable word order, for the three stages of the experiment, for sentences involving adjectives, numerals, or taken across the whole language. The dashed lines indicate the expected entropy of the training language for sentences involving adjectives and numerals separately; the solid line (at entropy = 1) gives the entropy of word order in the target language as a whole. Error bars indicate 95% CIs.

guage as a whole. The entropy of word order use for a participant is given by

$$Entropy = -\sum P(i) log_2 P(i)$$

where the sum is over the two possible word orders, Mod-N and N-Mod and $P(i)$ is the frequency of word order i in a participant's productions. Entropy measures how variable a participant's productions are: entropy of 0 corresponds to a participant who consistently uses a single word order, and entropy is at a maximum (entropy = 1) when both word orders are used equiprobably. Participants failed to reproduce the full variability of the target language, particularly for numerals; while the mean proportion of word order use across participants was close to the trained proportion (see above), the entropy analysis reveals that this masks a general tendency for participants to be more consistent within-category than their training data, and also more consistent across the two modifier types (i.e. more harmonic), as indicated by the whole-language entropy. Variability within the two modifier types and also across the whole language seems to reduce sharply in the second recall stage. The latter effect suggests that participants were preferentially using harmonic word orders (e.g. consistently noun-initial or noun-final) in the final recall test; this effect can in fact be seen in the individual data in Fig. 2, where several participants can be seen to substantially move word order for one modifier type towards the minority trained order, which corresponds to the majority trained order for the other modifier.

Statistical analysis confirmed these impressions: a regression predicting entropy from experimental stage (recall test 1 as reference level) and modifier type (adjective or numeral, centered) indicates a significant effect of modifier type ($\beta = 0.134$, $SE = 0.055$, $p = .014$), no difference in entropy between recall test 1 and interaction for either modifier type (as indicated by n.s. effect of experiment stage and n.s. interactions between modifier type and interaction stage: largest $\beta = -0.009$, $SE = 0.069$, $p = .901$), and significantly lower entropy at recall test 2 for both modifier types (as indicated by a significant difference in entropy between recall 1 and recall 2, $\beta = -0.183$, $SE = 0.060$, $p = .002$, and no interaction with modifier type, $\beta = -0.011$, $SE = 0.069$, $p = .867$).

Priming of word order use during interaction

Our central question in Experiment 1 concerns priming within and across the two construction types, Adj-N and Num-N; we focus on priming of the participant by Glermi, although analyses looking at priming from a participant's own previous productions produces a similar pattern of effects. Fig. 6 shows how the proportion of Mod-N orders produced by participants was influenced by the word order used by Glermi in the immediately preceding trial, for consecutive trials featuring both the same modifier type (e.g. where both Glermi and the participant used an adjective) and mismatched modifier types (e.g. Glermi produced an adjective, the participant produced a numeral). As well as showing the tendency to use N-Mod order in Adjective trials and a strong preference for Mod-N order in Numeral trials, this figure strongly suggests priming of word order on trials involving a matching modifier type (the participant's tendency to use Mod-N order was modulated by Glermi's word order when the modifier type involved was the same) but not across modifier types (the participant's tendency to use Mod-N order was unaffected by Glermi's word order when the modifier type involved was different).

Statistical analysis confirmed these impressions. In order to simplify the analysis, we collapsed the four combinations of trial types shown in Fig. 6 into two, coding each trial for whether the modifier involved was of the same category or a different category to that produced by Glermi (see Fig. 7). A logit regression predicting the participant's word order (Mod-N order or not) based on Glermi's order (Mod-N or not) and trial combination (same category of modifier or not), with both predictors centered, showed an overall preference for Mod-N order (as indicated by a significant intercept: $\beta = 0.699$, $SE = 0.339$, $p = .039$) and a significant interaction between Glermi's order and the trial combination ($\beta = 2.864$, $SE = 0.405$, $p < .001$; the independent effects of Glermi's order and the trial combination were not significant, largest $\beta = -0.173$, $SE = 0.201$, $p = .390$): participants were more likely to use Mod-N order themselves when Glermi had just done so *and* they were producing a modifier of the same category.

Finally, in order to establish whether the influence of a participant's partner's previous word order genuinely represents *structural* priming, i.e. copying of word order independent of the lexical items involved, we conducted a further regression analysis including as a predictor the lexical overlap (i.e. the number of words shared) between the focal utterance and Glermi's previous production, and the interactions between this predictor, Glermi's previous production, and trial combination, using 0 overlap as the model intercept. This model indicated a significant boost in priming from lexical overlap ($\beta = 1.725$, $SE = 0.773$, $p = .026$), i.e. a significant lexical boost, but, crucially,

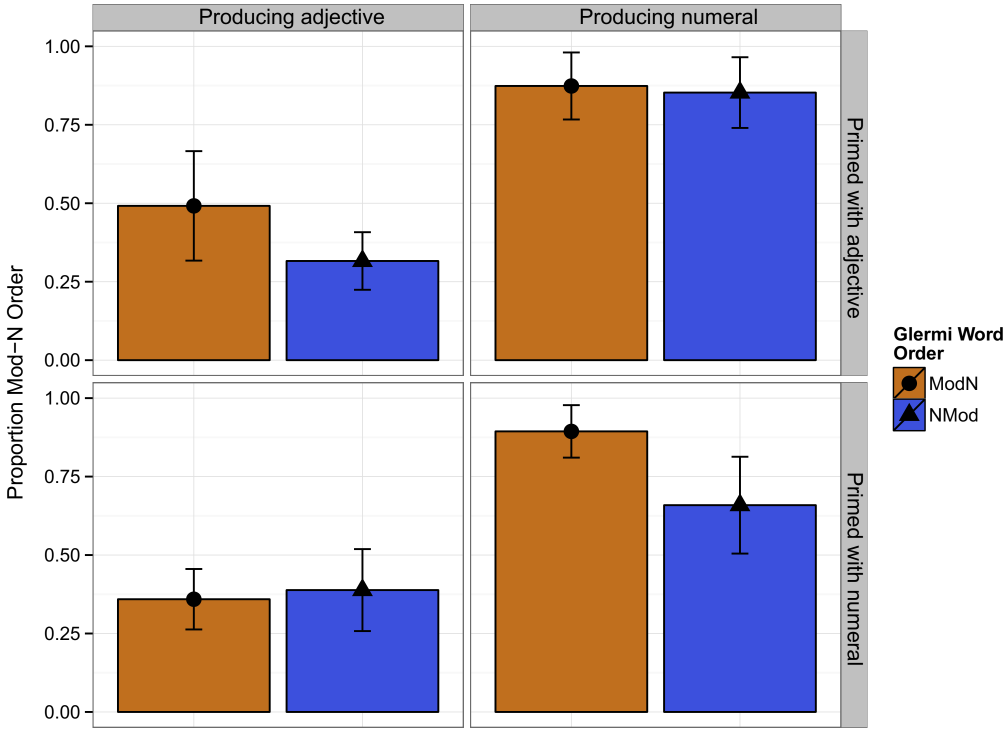

Fig. 6. Proportion of sentences produced in Mod-N order during interaction, broken down by the participant's modifier type, Glermi's immediately preceding word order, and Glermi's modifier type. Error bars indicate 95% CIs.

retained the interaction between the partner's previous production and trial combination indicative of within-category structural priming even on trials where there was no lexical overlap with Glermi's previous production ($\beta = 2.069$, $SE = 0.512$, $p < .001$).

Experiment 1 Summary

The results from Experiment 1 show that (1) participants reproduced the variation in their input language, using the word orders in roughly the same proportion as they were trained, although they were more successful in doing so for adjectives than numerals; (2) participants' choice of word order was modulated by Glermi's previous word order as long as the modifier was of the same type (noun or adjective); (3) this effect was genuinely structural, in that it was not restricted to trials in which lexical items recurred; (4) participants showed evidence of regularisation – they produced lower entropy output in the final recall test, relative to their input and to their behavior in recall test 1 and the interactive stage. This suggests that communicative interaction may indeed have a regularising effect on unpredictable variation, and provides evidence that language use itself plays a role in removing unpredictable variation from linguistic systems. However, the effect of structural priming during interaction works against regularisation in this experiment, since copying Glermi encourages variability – this suggests that the regularisation seen here is due to some other aspect of interaction, a point which we return to below.

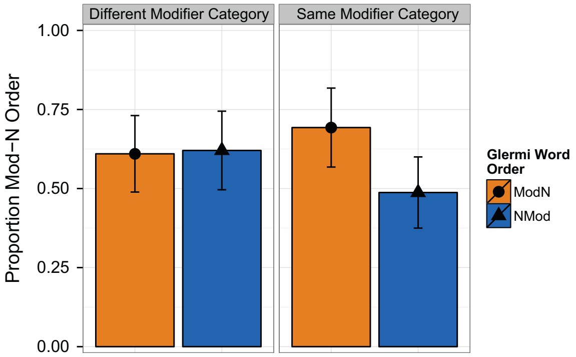

Fig. 7. Proportion of sentences produced in Mod-N order during interaction, broken down by Glermi's immediately preceding word order, and the match between the categories of the participant's modifier and Glermi's modifier. Error bars indicate 95% CIs.

We saw regularisation for both phrase types in isolation (i.e. participants became more consistent in their placement of adjectives relative to nouns, and in their placement of numerals relative to nouns). Interestingly, we also saw evidence of regularisation across the different phrase types, so that the final languages tended to have similar word order usage across modifier types (i.e. were more harmonic). This is surprising given that the specific language we employed here was one where Culbertson et al. (2012) did not see a preference for harmony. This suggests the possibility that communicative interaction provides additional pressures which may lead to the preponderance of harmonic languages, i.e. pressures beyond those exerted by individual learners (as seen elsewhere in Culbertson et al.'s data). However at present we feel that this result and interpretation should be treated with caution given that we did not see cross-category structural priming, which suggests that participants were not predisposed to generalize at this level.

One other unexpected aspect of our results was that, although for adjective phrases regularisation occurred equally around both possible word orders (i.e. entropy was reduced either because participants boosted the N-Adj order to make it more frequent than the Adj-N order, or vice versa), regularisation for numeral phrases was more often towards Num-N ordering. Culbertson et al. did not find the same bias in the equivalent condition in their data, although they did find a similar preference for Num-N ordering in another non-harmonic language type (Noun-Num and Adj-Noun). Although we cannot rule out some influence of English in our data, we did see regularisation towards the alternative N-Num order in several participants, and in fact, as Fig. 3 shows, we saw approximately equal numbers of participants regularising on the Num-N, Adj-N and the (less English-like) N-Num, N-Adj orders, with most participants producing a more regular form of the trained (Num-N, N-Adj) language. This suggests that biases from English are not driving regularisation in our data.

The presence of structural priming in our experiment has implications for artificial language learning, and may also speak to questions which have been of interest in the priming literature. First, it is important that we demonstrate true structural priming: our participants showed this effect even in the absence of shared lexical items. This suggests the presence of representations at the category level, even in these rapidly learnt miniature languages. This behavior is in line with what we see in experiments with natural languages, both with adult speakers and with young children whose language is still developing. The presence of these structural priming effects thus helps to validate the use of artificial language learning paradigms in investigating language learning and processing.

Returning to the relationship between structural priming and regularisation, as discussed above, it is important to realize that structural priming in Experiment 1 should work against regularisation, since Glermi's output reinforced the variability of word order. An important question, therefore, is what aspect of the experimental procedures led to the reduction of unpredictable variation, as seen in the reduction of entropy from recall 1 to recall 2. We suggest that use of the artificial language for communication, even with a non-human partner, changed participants' expectations about or recollection of the variability of that language. In particular, emphasizing the communicative use of the miniature language may have highlighted for our participants several counter-functional aspects of unpredictable variability discussed above, e.g. that a difference in form is usually taken (by humans) to indicate a difference in meaning and that frequently alternating word order requires breaking the current communicative convention established with one's interlocutor. Under this account, although regularisation was hindered by the low-level moment-to-moment priming that took place during interaction, once this was lifted in post-interaction recall, the regularising influence of interaction was revealed.

If we are correct that regularisation occurs because participants were (explicitly or implicitly) shaping their language for communicative purposes, we might expect that we would see even greater regularisation when participants believe themselves to be interacting with another human. Recall that Smith et al. (2014) and Fehér et al. (in preparation) found that pairs of humans communicating using variable artificial languages converged on languages with reduced linguistic variability. It is possible that the regularisation seen in this situation emerges, at least in part, from the participants' bias to avoid producing counter-functional variation for their human partner. On the other hand, regularisation in these human to human interactions is likely also affected by the possibility of reciprocal priming, which can allow reinforcement of one variant and convergence on a more regular system. Experiment 2 explores the roles of these different processes, using a new artificial language, by manipulating both participants' communication partner and their beliefs about their communication partner.

Experiment 2

In Experiment 1, we demonstrated two main results: category-specific structural priming and regularisation of variable word order following interaction. In Experiment 2, we extend our investigation into dyadic interaction between humans. In particular, we are interested in whether the *reciprocal* priming which occurs during interaction between two human interlocutors (A primes B who in turn primes A who in turn primes B ...) will result in stronger regularisation effects during interaction. Importantly, this reciprocal priming was not possible in Experiment 1: while our participants were primed by Glermi, Glermi's 'choice' of word order was unaffected by the participant's behavior and remained variable. In Experiment 2, we consider a more naturalistic situation in which two human participants interact with each other, and look for evidence of reciprocal priming and whether the amount of regularisation depends on the amount of priming. However, if we do see more regularisation in human-human interaction than in human-computer interaction, this may not necessarily result only from reciprocal priming, but also from a change in the participants' beliefs about their communicative partner. In Experiment 1, we argued that the regularisation seen at recall 2 resulted from strategic considerations (implicit or explicit) regarding the counter-functional properties of being unpredictably variable during communication. If this explanation is correct, it seems reasonable that participants will evidence even more regularisation for the benefit of a human interlocutor. To pull apart these different explanations, we include an additional condition in which participants interact with a computer but believe they are interacting with a computer.

Experiment 2 therefore features three conditions which allow us to manipulate the composition of dyads and the belief of our participants about their interlocutor. As in Experiment 1, we include a condition (the *Single* condition) where a participant knowingly interacts with a computer partner. Following Smith et al. (2014) and Fehér et al. (in preparation), we include a second condition (the *Dyad* condition) where two participants interact using the artificial language, in an approximation of naturalistic human-human interaction. In this condition we expect to see reciprocal priming, high levels of convergence in dyads (they will come to use the available word orders with similar frequencies), and regularisation (as a result of reciprocal priming leading participants to converge on regular languages, and/or strategic reduction in variability during interaction). Finally, we include an additional condition, the *Pseudodyad*, where we manipulate participants' beliefs about their communication partner: we tell them that they will interact with a human when, in reality, they will interact with a computer (i.e. this condition matches the human-human condition in terms of participants' beliefs, but the human-computer condition in the way their partner behaves during communication). Our main purpose in including the Pseudodyad condition is to explore the situation where there is no opportunity for regularisation to occur via reciprocal priming (because the computer partner cannot be primed and remains resolutely variable) and thus to isolate the role of participants' "strategic" preferences: if we see more regularisation in this condition than in the Singles condition, that would indicate that participants are regularising for the benefit of a (supposed) human interlocutor. Note also that, as in Experiment 1, in the conditions involving interaction with a computer partner, any bias for regularisation will have to outweigh the influence of priming from the partner's variable production.

In Experiment 2 we explore the relationship between priming and levels of representation in a new linguistic domain, moving to a paradigm in which, following Wonnacott et al. (2008), we train participants on a language with two synonymous word orders. The two forms we adopt are akin to the English Prepositional Object ["give the key to the man"] versus Double Object ["give the man the key"] constructions. Children's usage of verb-argument constructions has been shown to be sensitive to cross-verb structural priming, showing that they have abstract, verb-general construction representations (Messenger, Branigan, McLean, & Sorace, 2012; Peter et al., 2015; Rowland et al., 2012). It is therefore of interest to see whether participants will show similar evidence of abstract representations in their use of a newly learnt artificial language. Wonnacott et al. (2008) showed that adult participants learning a similar miniature language track the statistics of construction use not only at the level of specific verbs but also across verbs. This strongly suggests the presence of verb-general representations, but priming would provide a further test of how participants access these representations in production. We expect that people will generalize and therefore show priming across different verbs, although a lexical boost (stronger priming between sentences sharing a verb) is also in line with previous studies (Branigan et al., 2000; Pickering & Branigan, 1998).

Method

Participants

Fifty-two participants (14 males, 38 females, mean age 22.9) were recruited from the University of Edinburgh's Student and Graduate Employment service to take part in a miniature language communication experiment. Participants were paid £10 for their participation. Participants were randomly assigned to one of three conditions: 20 participants were run in the Dyad condition, 16 in the Pseudodyad condition, and 16 in the Single condition.

Procedure

Participants were seated in isolation in sound-attenuated booths, and worked through a computer program (written in Python using Psychopy: Peirce, 2007) which presented and tested them on an artificial language (see Fig. 8), and then allowed them to use that language to communicate remotely with their partner. As in Experiment 1, the language was text-based: participants observed pictures, videos and text displayed on the screen and entered their responses by keyboard.

Participants progressed through a six-stage training and testing regime, broadly similar to that used in Experiment 1.

Stage 1) Noun training: Participants viewed pictures of four puppet animals (bee, elephant, giraffe, lion) along with nonsense nouns which were transparently related to their associated referent animal (*buzzo*, *trunko*, *necko* and *roaro*). Each presentation lasted 5 s: for the first 2.5 s the picture of the animal was shown alone, then the nouns was presented alongside the figure for 2.5 s. Participants received 4 blocks of training, each consisting of one presentation of each noun in random order.[5]
Stage 2) Vocabulary testing: Participants were presented with a picture of an animal (each once, in random order), without accompanying text, and were asked to provide the appropriate label.
Stage 3) Sentence training: Participants were exposed to video clips paired with sentences describing those clips. Each clip showed a pair of puppet animals, with one animal performing one of four possible actions on the other: kissing, hugging, ramming with the head, rocking in the arms.[6] Each trial lasted 9 s: participants viewed the video clip once, ending in a freeze-frame bringing the total duration of the viewing up to 4.5 s, then the sentence was presented alongside a second viewing of the video clip plus freeze frame, lasting a further 4.5 s. The description accompanying each clip consisted of a nonsense verb (*smusa* for kiss, *ooshra* for hug, *wopla* for ram, *weewa* for rock), two nouns naming

[5] Presentation order for the two members of a pair was randomised independently throughout training and individual testing. In order to keep the participants roughly synchronized, participants were only allowed to progress to the next block of training/testing when their partner was also ready to begin the corresponding block.

[6] Every time a clip was played, it was either played as recorded (agents on the left) or mirrored (placing the agent on the right); this randomisation was carried out independently in every trial (so repeat viewings of a clip during training could differ in mirroring), and also independently for director and matcher during interaction, such that the director and matcher's clip would differ on mirroring on half of all interaction trials.

$$S \rightarrow V\ N_{Agent}\ N_{Patient} \quad (p = 0.5)$$
$$S \rightarrow V\ N_{Patient}\ tid\ N_{Agent} \quad (p = 0.5)$$

$$N \rightarrow \{\text{buzzo, trunko, necko, roaro}\}$$
$$V \rightarrow \{\text{smusa, ooshra, wopla, weewa}\}$$

Fig. 8. The grammar of the target language in Experiment 2. The language has two equiprobable word orders (VAP and VPpA), both of which are verb-initial but which differ in the order in which the agent and patient of the verb are expressed, with one of the orders being distinguished by the presence of a particle *tid*.

the participant animals, and (sometimes) a particle *tid*. Sentences were presented in one of two word orders (see Fig. 8 and also below): Verb – Agent Noun – Patient Noun (VAP), or Verb – Patient Noun – particle – Agent Noun (VPpA). For example, a clip in which a lion rammed a bee might be described as *wopla roaro buzzo*, or *wopla buzzo tid roaro*. Each of the 48 possible clips (4 agents × 3 patients × 4 actions) was presented three times (in three blocks, order randomised within blocks).

Stage 4) Individual recall test 1: Participants viewed the same 48 clips (order randomised) without accompanying text and were asked to enter the appropriate sentence. The video clip played on a loop while the participant typed their response, each completion of the clip being followed by 0.5 s of freeze frame. Each of the 48 clips was presented once, order randomised.

Stage 5) Interactive testing: Participants played a director-matcher game in which they alternated describing clips for their partner, and selecting a clip based on their partner's description. When directing, participants were presented with a clip (drawn from the set of 48 possible clips) and prompted to type the description so their partner could identify it. This description was then passed to their partner,[7] who had to identify the correct scene (by key-press) from an array of 4 possible clips. These 4 clips played on a loop, with each clip being followed by at least 0.5 s of freeze-frame before re-playing, and shorter clips remaining on freeze-frame so that all 4 clips were synchronized to start at the same time. The clips the matcher had to choose between always included the target clip and a second clip featuring the same two animals and the same action, with participant roles reversed; the array also included two further clips which differed from the target in either the action involved or one of the two participant animals, selected such that the matcher could not identify the target clip with above-chance probability simply by studying the set of clips (see Table 1). Since the director could not know the contents of the matcher's array, successful communication could only be guaranteed by encoding in the description all the details of the scene (the identity of the agent, the patient, and the action). After each trial both participants received feedback (success or failure) and an updated score ("Score so far: X out of Y"). Participants played 96 such communication games, such that each participant directed once for each possible clip (order randomised, interaction split into two blocks separated by a self-paced break after 48 trials, a randomly selected member of the pair directing first in each block and the participants alternating roles for the remainder of the block).

Stage 6) Individual recall test 2: Finally, participants viewed the same 48 video clips, order randomised, without accompanying text, and were asked to enter the appropriate sentence "in the language you learnt," all other details of the re-testing stage being identical to the first individual testing stage.

Variable word order during training

As discussed above, participants encountered two word orders during training, VAP and VPpA, which differed in the order in which the nouns referring to the agent and patient appeared, and which also differed in the presence of a nonsense particle marking the patient-first word order. We designed the training data each participant saw such that they encountered each word order equally often. The assignment of word orders to clips was randomised for each participant (but remained fixed across blocks for a given participant, e.g. if the scene featuring a lion ramming a bee was described with order VAP on training block 1, it would be described with that order on blocks 2 and 3 also), subject to the constraint that every noun, verb and action occurred equally frequently with each order, and every combination of agent and verb, patient and verb, and agent and patient occurred equally often with each word order. This construction of the test set ensured that word order was maximally unpredictable in training, and not conditioned on any aspect of the scene being described.

Manipulation of partners

We manipulated whether participants interacted with another participant or a computer script (see details below) during the interactive stage of the experiment, and in the latter case, whether they believed that they were interacting with a human or a computer. In the *Dyad* condition, participants were truthfully briefed that they would be interacting with another participant. In the *Single* condition, participants were truthfully briefed that they would be interacting with the computer; this condition is similar to Experiment 1, where participants were interacting with Glermi (who they presumably realized was the computer rather than a genuine alien). Finally, in the *Pseudodyad* condition, participants were given identical instructions as the Dyad condition, but in fact interacted with the computer throughout. The Dyad and Pseudodyad conditions were therefore identical in the briefing and

[7] In fact, to prevent participants communicating using English or any system other than the language they were trained on, the closest sequence of legal words was passed to their partner. The string produced by the director was split into a sequence of words bounded by whitespace, then each of those words was compared against all 9 possible legal words (4 nouns, 4 verbs, 1 particle), and the actual typed word replaced by the closest legal word (as measured by Levenshtein string-edit distance). This correction procedure applied on a word-by-word basis, and left the word order produced by participants unchanged. The resulting string of legal words was transferred to the matcher. As an additional block to prevent participants using English word order, if the director produced a description for which the first (corrected) word was not a verb in the training language, they received an on-screen warning ("I'm sorry, that doesn't seem to be a description from the language you learnt! Please try again.") and restarted the trial.

Table 1
Example test arrays. The matcher array is generated randomly on every interaction, and each of these possible matcher arrays is equally likely to occur.

Director sees	lion rams bee			
Matcher sees	lion rams bee	bee rams lion	lion hugs bee	bee hugs lion
or	lion rams bee	bee rams lion	lion rams giraffe	giraffe rams lion
or	lion rams bee	bee rams lion	giraffe rams bee	bee rams giraffe

(based on post-experiment debrief) the participants' beliefs about who they were interacting with; the Pseudodyad and Single conditions were near-identical (see below) in how interaction proceeded. We only ran pseudodyads when we had pairs of participants in the lab simultaneously, in order to maintain the illusion that they were genuinely interacting.

In the Pseudodyad and Single conditions, a computer script (written in Python) took the place of another participant. This script worked through all the same stages as the participants, and interacted with the participants in the interactive stage of the experiment. When acting as director during interaction, the computer director used both word orders in equal proportions – prior to commencing interaction, we assigned a word order to all 48 clips for the computer director, subject to the same constraints used to construct training data (i.e. the variation in word order was truly unpredictable), then the computer director simply produced those word orders during interaction. When acting as matcher during interaction, the computer matcher considered the 8 possible legal descriptions for the 4 scenes in its matcher array (two descriptions, VAP and VPpA, for each scene, the computer's matcher array being generated in exactly the same way as a genuine participant's matcher array) and simply selected the clip whose description was closest (according to Levenshtein distance) to the description provided by the director. In the case of one or more descriptions being equally close, the computer matcher selected randomly among those candidates. Finally, in order to maintain the plausibility of the Pseudodyad condition, the computer in the Single and Pseudodyad conditions differed in the speed with which they provided their responses: the computer in the Single condition provided all responses immediately, the computer in the Pseudodyad condition took a plausible amount of time on all trials (7–11 s when producing a description, 3–8 s when matching). As a result, it was possible that the human participant would have to wait for the computer to catch up during sentence training, individual recall and so on, as well as having to wait for the computer to produce and respond during interaction trials.

Analysis

All analyses were carried out using logit mixed-effects regressions with maximal random effects structures where possible (random intercepts and slopes by participant; in Dyads, the random effects for participants were nested within pair, and models included random intercepts and slopes by pair). For Condition we used the Dyad condition as our reference category; unless otherwise stated, all predictors other than Condition were centered. For analyses dealing with word order, we automatically identified the word order of participants' corrected judgements (i.e. the descriptions provided by the participants, filtered through the spell-checking procedure described above): every production was categorized as VAP, VPpA, or Other, and Other trials were excluded from analysis.

As in Experiment 1, we looked for (1) regularisation (during recall 1, interaction, or recall 2), which would be evidenced by a reduction in variability (as shown by over-production of one word order, or a reduction in entropy), and (2) structural priming during interaction, i.e. the tendency to reuse the word order used by their interlocutor (human or computer) in the immediately preceding trial. We also looked for (3) convergence, which is an indicator of the similarity in language use between communicating partners. For two interlocutors (both human, or human plus computer), we simply measured the difference in their usage of the two possible word orders, with low difference indicating similar linguistic behavior and therefore high convergence.

Results

Communicative success

Performance during the communicative portion of the task was extremely high throughout, and varied relatively little across conditions (see Fig. 9). Logit regression with Block and Condition as fixed effects showed a significant effect of block ($\beta = 1.042$, $SE = 0.489$, $p = .033$), and significantly lower communicative success in Pseudodyads and Singles, relative to Dyads (Pseudodyads: $\beta = -0.909$, $SE = 0.454$, $p = .045$; Singles: $\beta = -1.326$, $SE = 0.444$, $p = .003$; the interactions between condition and block are n.s., largest $\beta = -0.372$, $SE = 0.580$, $p = .522$).

Word order use

Fig. 10 shows the full data for word order use across individual testing (Recall 1), interaction, and individual re-testing (Recall 2); Fig. 11 shows the data from participants in the Dyad condition, organized by pairs. These

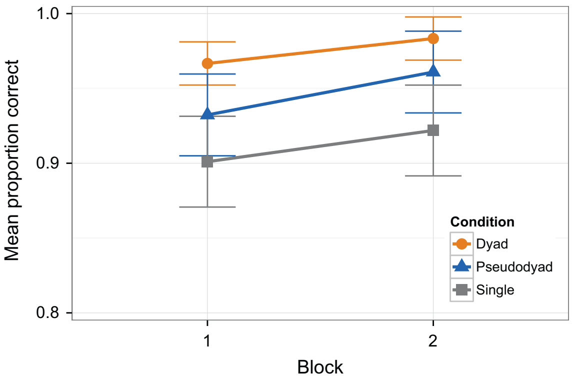

Fig. 9. Proportion of successful communication trials during the two blocks of interaction. Error bars indicate 95% CIs.

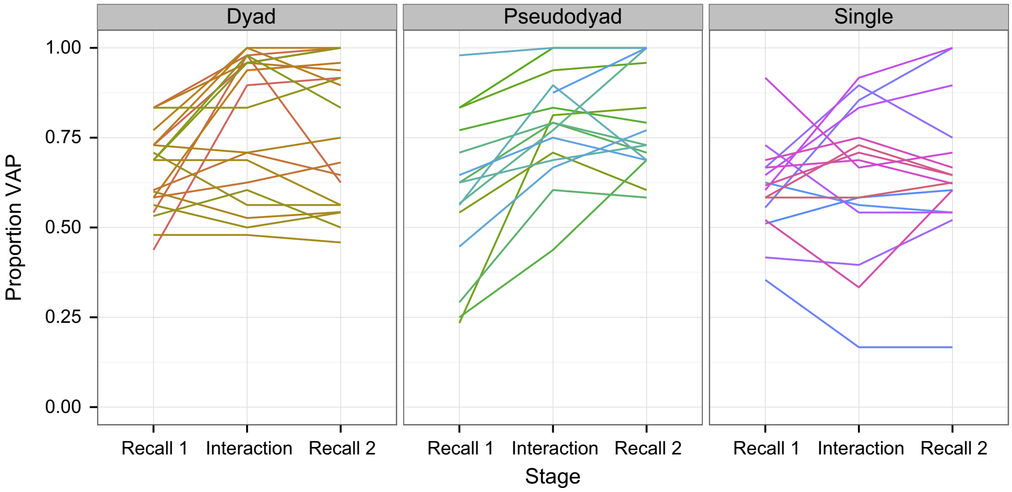

Fig. 10. Proportion of VAP sentences produced during recall test 1, interaction, and recall test 2. Each line represents a participant, participants who form a pair in the Dyad condition are represented by lines of the same color. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

figures indicate that our participants produced variable linguistic output during the individual recall test immediately following training, and were generally more variable than participants in Experiment 1. All participants who produced grammatical sentences during testing produced both word orders (one participant in the Pseudodyad condition produced only verbs during the first individual test, although subsequently produced grammatical sentences during interaction). However, in line with the results from Wonnacott et al. (2008), participants preferred the VAP word order: only a handful of participants produced VPpA on the majority of recall trials.[8]

A logit regression exploring the effect of Condition and test block (recall test 1, interaction, recall test 2, with recall test 1 as intercept) indicates that participants in all conditions produced VAP order more often than expected by chance in recall test 1 (as indicated by a significant model intercept, $\beta = 0.688$, $SE = 0.126$, $p < .001$, and no effect of Pseudodyad or Single condition: largest $\beta = -0.303$, $SE = 0.195$, $p = .120$). Participants in the Dyad and Pseudodyad conditions increased their proportion of VAP productions during interaction and at post-interaction recall (for Dyads, log-odds of VAP is higher at interaction and recall test 2: interaction stage $\beta = 1.164$, $SE = 0.317$, $p < .001$, recall test 2 $\beta = 0.901$, $SE = 0.339$, $p = .008$; Pseudodyads show the same increase in use of VAP during interaction and recall test 2, as indicated by n.s. stage × Pseudodyad interactions, largest $\beta = 0.420$, $SE = 0.448$, $p = .349$). Participants in the Single condition showed significantly less increase in use of VAP during interaction (significant interaction between Single condition and interaction test block: $\beta = -0.870$, $SE = 0.409$, $p = .033$; note the negative slope of approximately the same magnitude as the effect of interaction stage in the other two conditions), although their use of VAP was not significantly lower than that seen in Dyads during recall test 2 ($\beta = -0.460$, $SE = 0.438$, $p = .293$). These effects are visible in Fig. 10, but is illustrated for clarity in Fig. 12. Note that participants in the Pseudodyad and Single conditions had identical input from their 'partner' during interaction, suggesting that this difference in their use of VAP in this stage of the experiment is driven by differences in the participants' own behavior during interaction, or their perception of their partners' behavior, or both. These results for word order use in the Single condition parallel those seen in the Experiment 1 – the average word order use of participants who interact with a computer partner remains relatively constant across recall test 1 and interaction.

These same results are replicated in an analysis of the entropy of participants' productions (not shown):

[8] In their training data, participants saw each scene labelled in a consistent manner across all three blocks – i.e. if the scene involving a bee kissing a giraffe was presented with VAP order in block 1, it would also be presented with VAP order in blocks 2 and 3. This raises the possibility that participants might condition (or partially condition) their word order use on this regularity in their training data. To check for this scene-specific conditioning of word order, we calculated for each participant how often the word order that participant produced (during recall 1, interaction or recall 2) for a given scene matched the training order they saw for that scene – i.e. returning to the example above, if the participant produced VAP for "bee kisses giraffe", that would be scored as a match. We then compared the observed number of matches to the number of matches obtained in 10,000 randomisations of the participant's productions – this gives us a distribution of matches between training and produced word orders which exactly matches the distribution of word orders produced by the participant, but which is by design unconditioned on scene-specific order encountered during training. We can then extract the probability of the veridical number of matches being obtained by unconditioned use of the word orders from this distribution – if less than 5% of the random distribution shows an equal or greater number of matches, then for that participant we can reject the null hypothesis that they are producing word orders unconditioned on the scene-specific training order at $p < .05$, one-tailed. We ran this statistic for all 52 participants at all three stages of the experiment (recall 1, interaction, recall 2). We could only (tentatively) reject the null hypothesis of unconditioned word order use for a single participant in recall 1, at $p = .0341$; no other participant fell below $p = .05$ at recall 1, and no participant fell below $p = .05$ for any other stage; indeed, the vast majority of observed p values were far larger than this. In other words, there is strong evidence that our participants were not conditioning their word order use in a scene-specific way, despite the consistent by-scene word order they encountered during training.

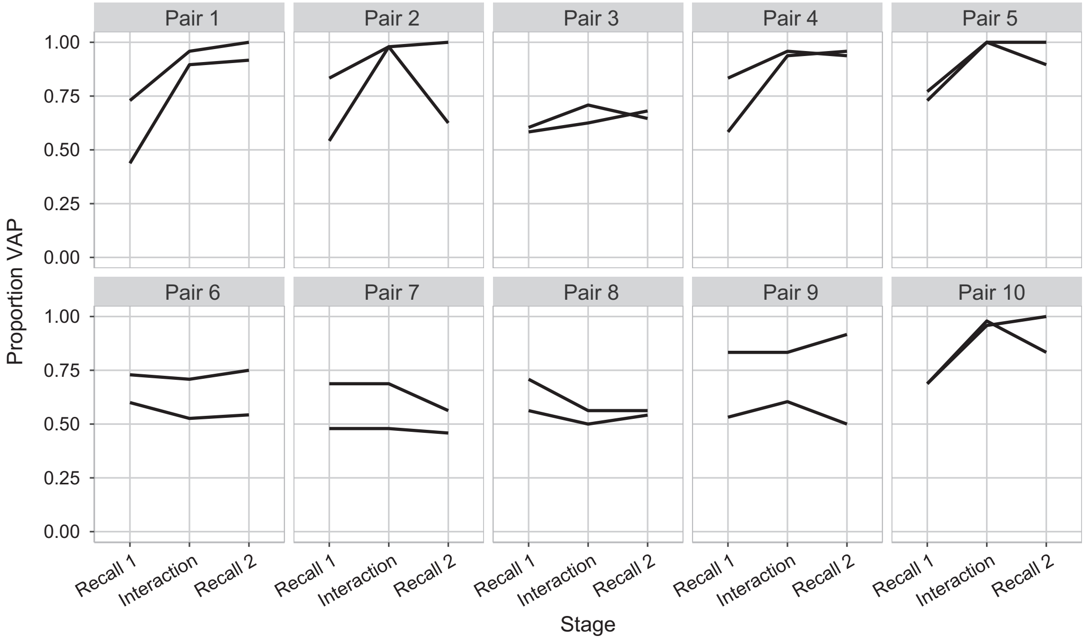

Fig. 11. Proportion of VAP sentences produced during recall test 1, interaction, and recall test 2 by the communicating pairs in the Dyad condition. Each subplot shows a pair, and each line represents a participant.

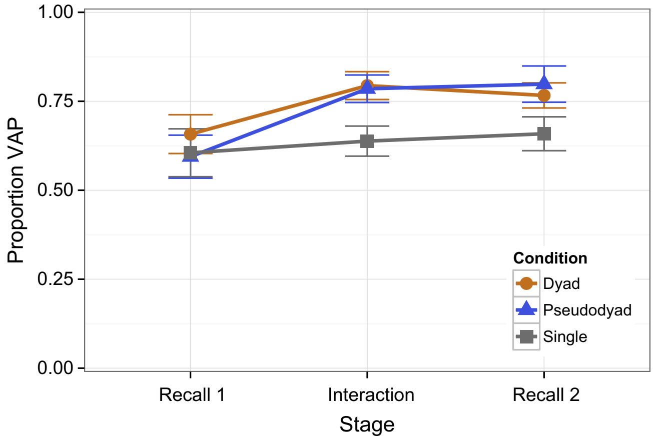

Fig. 12. Mean proportion of VAP sentences produced during recall test 1, interaction, and recall test 2. Error bars indicate 95% CIs.

participants in the Dyad and Pseudodyad conditions reduce the variability of their productions during interaction and at post-interaction recall, while entropy in the Single condition remains relatively high and constant across the experiment, although (as in the production data) patterns more closely with Dyads and Pseudodyads at recall 2 (as evidenced by a significant interaction between interaction stage and the Singles condition, but no interaction between recall test 2 and the Singles condition); the dramatic drop-off in entropy seen in Experiment 1 at recall test 2 is absent.

While dyads and pseudodyads were therefore closely matched in their average use of VAP order, closer inspection of the distribution of VAP responses (see Fig. 13) reveals a subtle difference: while participants in the pseudodyads are relatively tightly clustered around the mean proportion of VAP productions, dyads show a much more widely dispersed, somewhat bimodal distribution, reflecting the fact (as can also be seen in Fig. 11) that approximately half of our dyads converge on highly regular systems of word order use, using VAP order almost exclusively, a phenomenon which is relatively rare in pseudodyads. This difference in the spread of data between these two conditions is confirmed by Levene's test ($F(1,34) = 5.715$, $p = .0225$).

Priming of word order use during interaction

Fig. 14 shows, for each condition and separated by block of interaction, the proportion of trials on which participants produced sentences according to VAP order, based

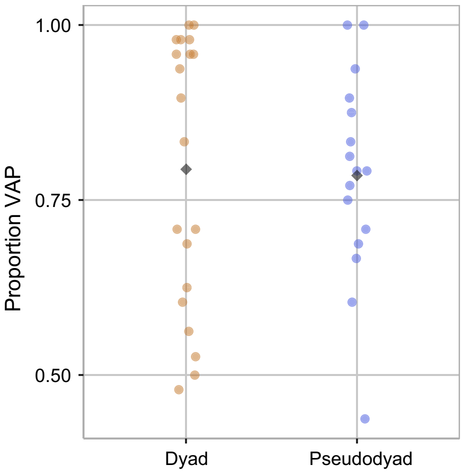

Fig. 13. Proportion of VAP sentences produced during interaction by participants in the Dyad and Pseudodyad conditions. Each point represents a participant (points are jittered horizontally to prevent overplotting), and the mean for each condition is given by a black diamond.

on their partner's word order in the immediately preceding trial. The data strongly indicate priming in all three conditions: participants are more likely to employ VAP order if their partner produced a sentence in VAP order.

A logit regression predicting word order (VAP or VPpA) based on three predictors – partner's previously produced word order, (i.e. the word order the participant received when last acting as matcher, NA for their first matcher trial), block of interaction and condition (with Dyad as reference level) – confirms this impression. As previously noted, Dyads and Pseudodyads over-use VAP order during interaction, significantly more so than Singles (as indicated by a significant model intercept, indicating log odds of using VAP significantly higher than zero, no effect of Pseudodyad condition but a significant effect of Single condition: intercept $\beta = 1.817$, $SE = 0.390$, $p < .001$; n.s. effect of Pseudodyads, $\beta = 0.046$, $SE = 0.512$, $p = .929$; significant negative effect of Single, $\beta = -1.004$, $SE = 0.504$, $p = .046$). There is also a tendency to use VAP order more in the second block of interaction in all conditions (the effect of block is significant for dyads, $\beta = 0.707$, $SE = 0.250$, $p = .005$; no significant interaction with the other conditions, largest $\beta = -0.445$, $SE = 0.328$, $p = .176$). Importantly, there is a significant effect of partner's previous word order in Dyads ($\beta = 1.307$, $SE = 0.353$, $p < .001$), which is also seen in the other two conditions (as indicated by the lack of significant interactions with Condition, largest $\beta = -0.580$, $SE = 0.445$, $p = .192$): in all three conditions, participants were more likely to use VAP order if their partner had used VAP on the preceding trial.

Finally, in order to establish whether the influence of a participant's partner's previous word order genuinely represents *structural* priming, i.e. copying of word order independent of the lexical items involved, we conducted a further regression analysis including as a predictor whether or not the verb matched that in the partner's previous production. Since the effects of lexical overlap should be restricted to the priming utterance, we re-ran the regression analysis outlined above, adding as a predictor the interaction between the partner's previous production

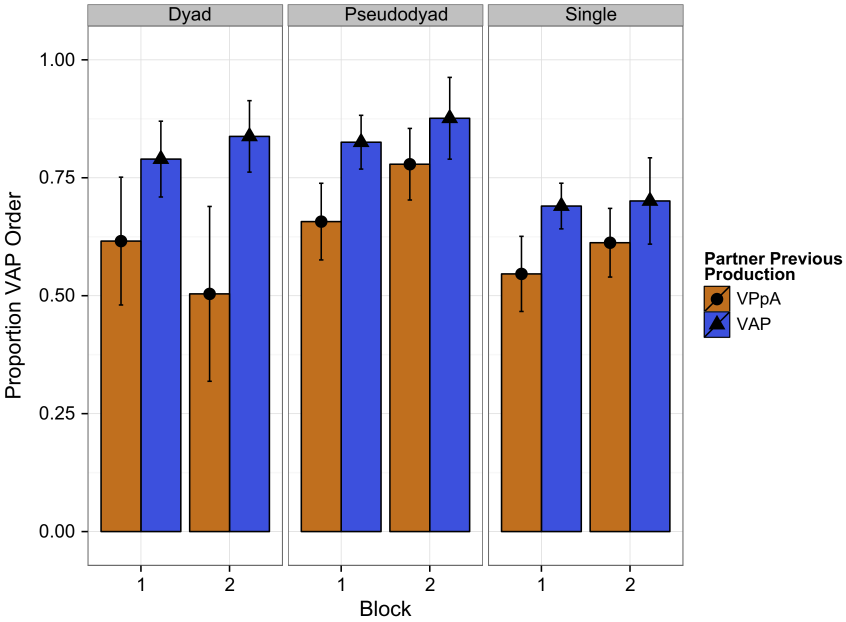

Fig. 14. Proportion of sentences produced using VAP order during interaction, broken down by condition, interaction block, and partner's immediately preceding word order. Participants across all blocks and all conditions tend to use VAP order more often when their partner has just used VAP order. Error bars indicate 95% CIs.

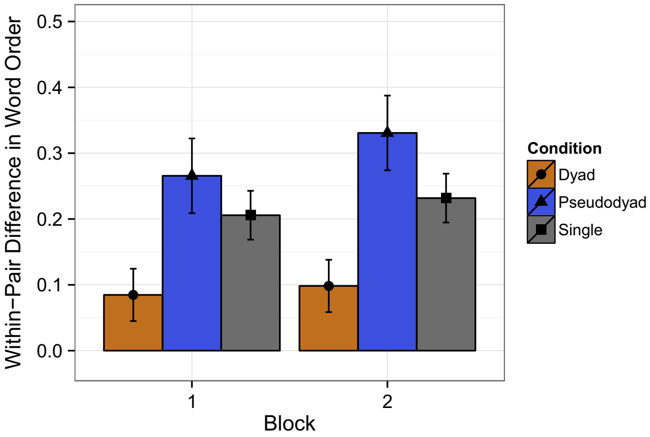

Fig. 15. Within-pair difference in frequency of use of VAP order, across the two blocks of interaction in all three conditions. Low within-pair difference corresponds to convergence in word order use. Error bars indicate 95% CIs.

and the lexical overlap with that production, taking 0 overlap as the model intercept. This more complex model[9] retained the significant effect of the partner's previous production, indicating priming ($\beta = 0.470$, $SE = 0.213$, $p = .027$: recall that this is reported at the reference level for verb overlap, i.e. verb-different trials, and thus indicates that structural priming occurs across verbs). Moreover, adding this interaction did not significantly improve model fit ($\chi^2(1) = 0.374$, $p = .541$), indicating the absence of a lexical boost in this experiment but again indicating that priming is not restricted to sentences sharing the same verb; in conjunction with the results of Experiment 1, this therefore gives us some confidence that the priming we see is genuinely structural.

Convergence within pairs

Finally, we investigated the degree of convergence within pairs: for each pair (either human-human or human-computer) we measured the magnitude of the difference in their frequency of use of VAP order (e.g. in a pair where one individual used VAP order on 75% of trials and their partner used it on 85% of trials, the within-pair difference would be 10% or 0.1). This measure of within-pair convergence is shown in Fig. 15. As we have shown in the previous sections of the results, participants in Dyads used the two word orders with the same mean frequency as participants in the Pseudodyad condition, produced similarly variable word order distributions (as measured by entropy), and were primed to approximately the same degree; nonetheless, they were substantially more similar to their human interlocutors than participants in the Pseudodyad and Single conditions were to their computer partner. This is confirmed by a regression analysis predicting within-pair difference on the basis of Condition (Dyads as reference level) and Block of Interaction. Level of within-pair alignment was constant across blocks (n.s. effect of block, $\beta = 0.014$, $SE = 0.037$, $p = .714$, and no interactions involving block, largest $\beta = 0.051$, $SE = 0.047$, $p = .278$), suggesting that Dyads converged with their partners very rapidly; Dyads were significantly more converged than either Pseudodyads ($\beta = 0.207$, $SE = 0.046$, $p < .001$) or Singles ($\beta = 0.127$, $SE = 0.046$, $p = .005$).

Summary of Experiment 2

In Experiment 2 we found that (1) although participants had a preference for one of the two word orders (VAP), they reproduced the variability of word orders in their input language; (2) participants in the Dyad and Pseudodyad conditions regularised the language significantly during interaction, and significantly more than participants in the Single condition; (3) the output participants produced in the final recall test was more regular than the output they produced in the recall test immediately following training, paralleling what we saw in Experiment 1 (although this effect was the least pronounced for the condition which most closely matched Experiment 1, i.e. the Singles condition) and showing that interaction can have both immediate (during interaction) and lasting (post-interaction) effects on linguistic variation; (4) structural priming occurred in all conditions, with participants in all three conditions being more likely to use the VAP order if their partner used it, even when the constructions contained different verbs; (5) despite being closely matched to pseudodyads on nearly all measures (i.e. showing approximately the same mean marker use during interaction and the same degree of priming), participants in dyads closely converged with their (human) interlocutors, whereas participants in the Pseudodyad (and Singles) condition failed to closely mirror the word order use of their highly variable partners; (6) despite being closely matched to pseudodyads on mean marker use during interaction, participants in dyads showed a stronger tendency to produce highly regular output during interaction.

In summary, our results show that interaction induces regularisation, particularly when participants believe that

[9] Getting the model with full random effect structure plus this additional interaction to converge proved difficult: consequently, here we compare models featuring only random intercepts, for the models including and lacking the interaction with lexical overlap.

they are interacting with another human. We saw similar levels of structural priming in all conditions but we only saw convergence on a shared system of word order use in genuine dyads, featuring human-human interaction. Languages entirely lacking variability were more likely to develop in human-human interaction, where reciprocal priming was possible. Note that convergence was possible even in human-computer interaction, though it would involve the participant converging on the highly variable system used by the computer. Instead of doing this, however, the participants in pseudodyads produced output which was less variable than that produced by their computer interlocutor. This suggests that linguistic convergence and structural priming (local, moment-to-moment alignment) can be dissociated at least to a certain degree.

The fact that we saw equivalent levels of regularisation in dyads and pseudodyads suggests that the pressures for regularisation which emerge in interaction do not depend solely on reciprocal priming. Regularisation was stronger in dyads and pseudodyads than in the Singles condition (i.e. we saw strong regularisation in interaction whenever participants believed they were interacting with another participant, even if they were in fact interacting with a consistently variable computer partner). This speaks to the mechanisms driving regularisation: interaction highlights the counter-functionality of unpredictable variation, leading participants to shape their own output to be more regular (and thus more functional). This occurred most strongly when participants believed their interlocutor to be human. However, there was a subtle difference in the types of systems produced in dyads and pseudodyads, in that the most regular systems we saw (lacking word order variability) tended to be produced during human-human interaction, where reciprocal priming was possible. Our data therefore suggest that reciprocal priming might act in concert with strategic avoidance of variation during interaction and learning biases against unpredictable variation to remove unpredictable linguistic variation when it is present.

General discussion

Regularisation of structural variability

We presented two experiments using artificial languages that exhibited unpredictable, probabilistic variation in word order to investigate regularisation and priming in three different communicative situations: when participants interacted with a computer, in natural human-human dyads and in situations when participants believed they were interacting with a human but actually interacted with a computer. Participants in both experiments acquired and reproduced the syntactic variation of the target language during an initial recall test. Furthermore, participants who knowingly interacted with a computer partner produced variable output during interaction and final recall, although their languages became more regular during post-interaction recall (as seen in Experiment 1 and the Singles condition of Experiment 2). Those participants in Experiment 2 who interacted with other participants or believed they were doing so (Dyad and Pseudodyad Conditions) regularised their languages during interaction in the direction of their initial word order preference – they were more likely to use one of the two variants (word orders) consistently across their productions.

As well as seeing effects on word order use during interaction, we observed lasting effects in post-interaction recall: in Experiment 1 we saw an increase in regularity and harmony in post-interaction recall, whereas in Experiment 2 the production of the dominant VAP order was higher at post-interaction than pre-interaction recall in all conditions, but particularly in dyads and pseudodyads. Both the effects seen during interaction and at post-interaction recall are of potential relevance in explaining how interaction might drive language change and ultimately shape the structure of natural languages. The lasting effects of interaction on individuals' internal grammar, which we assume were captured by the post-interaction recall test in our experiments, will influence the subsequent linguistic behavior of that individual, and therefore the linguistic data which forms the basis for learning in other members of the population. However, the changes that happen during interaction play an equally important role in shaping the linguistic data that new learners learn from: since children acquire language from linguistic data which is produced during communicative interaction, processes which systematically shape that data (e.g. the tendency to under-represent variability during interaction) will influence the data that children learn from, and ultimately the structure of the language. In other words, while we do see lasting effects of interaction on the grammars of individuals across all conditions in both of our experiments, we believe these lasting effects at the individual level are not necessary for interaction to influence language design. An important area for future work is to combine these dyadic interaction paradigms with iterated learning (following e.g. Kirby et al., 2015; Winters, Kirby, & Smith, 2015) to explore how the changes introduced during interaction accumulate over transmission and diffuse through populations.

We found that regularisation was stronger when participants believed they were interacting with a human, even when they were actually interacting with a computer. This was not due to a difference in the strength of priming between the Pseudodyad and Single conditions (stronger priming from the computer would in fact act against regularisation), since structural priming was equivalent across those conditions. It also does not seem to be the result of convergence (which would again work against regularisation, since convergence would require participants to use the variants in the same maximally variable proportion as the computer), since this was only seen in genuine dyads. One possibility is that participants believe (explicitly or implicitly) that variability during communication will be potentially counter-functional, and that producing a more regular language will therefore facilitate communication with their human partner, an interpretation which is consistent with data from Perfors (2016). Interestingly, however, we do see dyads who converge on highly variable systems and nonetheless communicate extremely successfully, suggesting that this is not necessarily an accurate intuition. Future work will further probe the nature of participants' beliefs about their partners and the benefits of regularity in communication.

A surprising aspect of our results is that we saw approximately equal amounts of regularisation in dyads and pseudodyads, despite the fact that reciprocal priming was only possible in dyads, suggesting that reciprocal priming does not play a role in regularisation, at least in our data. However, only in dyads did we see large numbers of participants producing highly regular languages, lacking word order variation, indicating that reciprocal priming may be crucial in eliminating variability entirely. We also only saw high levels of within-pair convergence in dyads, suggesting that reciprocal priming may be an important mechanism by which shared regular language systems emerge. Note that this may be particularly important in situations where individuals are biased to regularise in different directions.

It is worth noting that our participants received only a short period of training on the target language prior to interaction – while this was sufficient for them to acquire the target language with reasonably high accuracy, it may be that their limited experience with the language prior to interaction means they were more susceptible to factors affecting their language use during interaction. Our prediction would be that increasing the amount of pre-interaction training would reduce the extent to which participants change their language use during and after interaction, but that the same pattern of results (priming and greater regularity during interaction, greater regularity during post-interaction recall) would hold under a wide range of circumstances. Further experimental work systematically manipulating amount of training would be required to test this prediction.

The tendency we saw in Experiment 2 to regularise on VAP order was likely mediated to some extent by the similarity between VAP order and English (the native language of our participants). Similarly, in Experiment 1 we observed less flexibility in Numeral-Noun ordering and more flexibility for the Adjective to come before or after the noun in participants' output languages, which mirrors the different tendencies of these two modifiers in English (Goldberg, 2013). However, our results cannot simply be explained by native language influence. As Fig. 3 clearly shows, participants in Experiment 1 were just as likely to regularise on the anti-English order of Noun-Num and Noun-Adj as the English order (3 individuals vs. 5). In addition, while we did see an overall preference for VAP in Experiment 2, half of the communicating pairs in the Dyad condition produced variable output, rather than regularising on the most English-like order. Other studies have shown that, given sufficient training on a statistically skewed language, the preference for VAP order in this paradigm can be overcome (Wonnacott et al., 2008, Experiment 3). A clear prediction of our work here is that in such circumstances pairs of interacting participants would align on highly regular VPpA order during interaction, rather than simply overriding the statistical properties of their shared experience with the artificial language and converging on the VAP order.

Structural priming in artificial languages

We found roughly equal amounts of priming across all conditions in our experiments: participants were more likely to repeat the immediately preceding word order choice of their partner regardless of the overall amount of regularisation and convergence (discussed below). This suggests that priming was largely automatic, since it was not mediated by participants' expectations of their interlocutor's knowledge or behavior. Contrary to our findings, Branigan et al. (2011) found stronger priming effects in human-computer interaction and even stronger priming when people communicated with computers perceived as less capable, which the authors attribute to communicative design. In our experiments people were instructed, taught and tested by the computer and they had no reason to doubt its capabilities or proficiency. The fact that we did not observe significant differences in the strength of priming between pseudodyads and singles is potentially informative, since participants who believed they were interacting with another participant (i.e. in the Pseudodyad condition) might reasonably have expected their partner to be less capable than the computer.

From the perspective of the priming literature, our findings regarding the 'lexical boost' (increased priming in the presence of shared lexical items) are potentially meaningful. The lexical boost has been found to be robustly present in natural language experiments involving adult language users, though not in young children (Peter et al., 2015; Rowland et al., 2012). The fact that we saw this boost in Experiment 1 might suggest that the lexical boost is a consequence of being adult, rather than the stage of language development (since our participants are all in the earliest stages of learning this language), which would be consistent with the hypothesis given by Peter et al. (2015), who argue that lexical boost may rely on an ability for explicit memory which is more developed in adults. However, despite the fact that adult learners in this paradigm are able to track verb-specific preferences for specific structures (Wonnacott et al., 2008), we found no evidence for a lexical boost in Experiment 2, which could be used to draw the opposite conclusion, namely that the absence of a lexical boost is characteristic of the early stages of acquisition. Since the data across our two experiments are therefore equivocal, we can draw no strong conclusions on this issue. However, artificial language learning provides a means to dissociate extensive linguistic experience from age, and future studies could therefore use these techniques to tease apart these two possible accounts of the lexical boost.

Our results also have implications for the relationship between behavioral convergence and alignment. As discussed above, structural priming resulted in linguistic convergence in human-human dyads, but not in pairs featuring human-computer interaction, despite the presence of priming in all three conditions. Pseudodyads and singles, although primed by their partner (the computer), did not align with its use of word orders (which would have meant producing highly variable output in the same proportions as the computer during interaction), and increased their regularity over the course of interaction. This suggests that priming is a low-level mechanism that occurs even in the absence of convergence, and although it may play an important part in communicative alignment (Pickering & Garrod, 2004), its presence does not inevitably lead to behavioral convergence.

Behavioral alignment may be affected by a number of factors: phonetic convergence has been shown to be influenced by the participant's gender and role in the interaction (Pardo, 2006), and their initial language distance (Kim, Horton, & Bradlow, 2011). In addition, phonetic and structural convergence is affected by speakers' social perception of their interlocutor either positively, causing alignment (Balcetis & Dale, 2005; Giles et al., 1973; Giles & Powesland, 1975), or negatively, resulting in anti-alignment (Balcetis & Dale, 2005; Bourhis et al., 1979; Doise et al., 1976). In our experiment, these factors were held constant, as people's knowledge of the language was equal and well-controlled (one of the major advantages of the artificial language paradigm) and participants did not receive social information about their interlocutors, although we cannot completely rule out that gender (which was not controlled for) and personality type (on which we did not collect data) had an influence on linguistic alignment. It is interesting that even though our participants knew very little about their interlocutors we saw evidence of convergence which was absent in human-computer interaction. We suggest that this was due to the possibility of reciprocal priming in this condition, whereby participants repeatedly primed each other in turn. Future research can further explore the automaticity of this process and whether it can be influenced by the social factors indicated by previous research. More generally, manipulating social cues during the learning (as in Fehér, Kirby, & Smith, 2014) and use (as in Kerr & Smith, 2016) of artificial languages is a potentially profitable avenue for future enquiry.

Category-specific and category-general priming

In Experiment 1, different levels of regularisation for different modifier types and the absence of cross-category priming suggest that participants failed to form a representation for "modifier" as a general category subsuming the adjective and numeral categories. At the same time, in the final recall tests, we did find evidence of regularisation occurring across the numeral and adjective categories, suggesting some evidence of a bias to form a category at this level. A possible explanation for why our participants failed to form a modifier category which was sufficiently robust to allow priming is that we trained participants on a non-harmonic language in which distributional information (numerals and adjectives tend to occur on opposite sides of the noun) strongly cued against a higher-order modifier category. A further test for this would be to train participants on harmonic languages and measure cross-category priming. Positional information has been shown to aid category learning in adults (e.g. Frigo & McDonald, 1998; Gomez & Gerken, 1999; Hudson Kam, 2009; Mintz, 2002; Smith, 1966), and in children (Geffen & Mintz, 2014), and promote the formation of grammatical categories in artificial languages as well (Reeder, Newport, & Aslin, 2009; Reeder et al., 2013), which strongly suggests that in the case of a harmonic language, participants would be able to form representations at higher levels and therefore exhibit priming across modifier types.

In Experiment 2, we moved to a new artificial paradigm involving verb argument structures. Here, there is a clear prediction that participants should show verb-general priming since Wonnacott et al. (2008) used a similar paradigm to demonstrate that participants can acquire both verb-specific and verb-general distributional information. We found evidence of verb-general priming in this experiment, providing further evidence that, like children acquiring a natural language, learners of these languages form robust and abstract generalisations.

Conclusions

Our results indicate that classic techniques from developmental and psycholinguistic traditions – artificial language learning, scripted and unscripted interaction – can profitably be combined, with substantial potential benefits to both fields. We have combined these techniques here to demonstrate structural priming in two artificial language learning paradigms, and to explore the effects of interaction on unpredictably variable linguistic systems. We found that, as with natural languages, structural priming was robustly present when participants produced the language in an interactive setting. In addition, interaction lead to a reduction in unpredictable linguistic variation, although this appeared to be mainly due to a tendency to avoid producing counter-functional unpredictable variation during interaction, rather than from reciprocal structural priming. Future work will establish whether reciprocal priming is nevertheless an important mechanism in allowing language users to arrive at shared regular systems. More generally, research on structural priming can benefit from the well-established advantages offered by artificial language learning with respect to experimental control over the nature of participants' pre-experiment exposure to the structure of interest. Furthermore, artificial language learning could allow the priming of structures not attested in participants' native languages to be explored. In return, priming offers researchers working with artificial languages a new and powerful tool to probe the representations underlying their participants' linguistic knowledge. Most importantly, these techniques together can be useful in exploring the communicative mechanisms leading to language change and, ultimately, to the universal properties natural languages exhibit.

Acknowledgments

This work was supported by the Economic and Social Research Council (Grant No. ES/K006339). We thank Jennifer Culbertson for providing code and stimuli for Experiment 1, and two anonymous reviewers plus our colleagues at the Centre for Language Evolution for helpful feedback on this work.

Please cite this article in press as: Fehér, O., et al. Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language* (2016), http://dx.doi.org/10.1016/j.jml.2016.06.002

References

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by human infants. *Psychological Science, 9*, 321–324.

Balcetis, E., & Dale, R. (2005). An exploration of social modulation of syntactic priming. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the Cognitive Science Society* (pp. 184–189). Mahwah: Lawrence Erlbaum.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General, 129*, 177–192.

Bourhis, R. Y., Giles, H., Leyens, J.-P., & Tajfel, H. (1979). Psycholinguistic distinctiveness: Language divergence in Belgium. In H. Giles & R. N. St. Clair (Eds.), *Language and social psychology* (pp. 158–185). Baltimore: University Park Press.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition, 75*, B13–B25.

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition, 104*, 163–197.

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

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

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*, 711–733.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2010). A vowel is a vowel: Generalizing newly learned phonotactic constraints to new contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 821–828.

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

Clark, E. (1988). On the logic of contrast. *Journal of Child Language, 15*, 317–335.

Coupland, N. (2010). Accommodation theory. In J. Jaspers, J. Verschueren, & J. Östman (Eds.), *Society and language use* (pp. 21–27). Amsterdam: Benjamins.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology, 1*, 42–45.

Croft, W. (2000). *Explaining language change: An evolutionary approach*. London: Longman.

Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition, 139*, 71–82.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition, 122*, 306–329.

Doise, W., Sinclair, A., & Bourhis, R. Y. (1976). Evaluation of accent convergence and divergence in cooperative and competitive intergroup situations. *British Journal of Social and Clinical Psychology, 15*, 247–252.

Fehér, O., Kirby, S., & Smith, K. (2014). Social influences on the regularization of unpredictable variation. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the Cognitive Science Society* (pp. 2187–2191). Austin, TX: Cognitive Science Society.

Fehér, O., Ritt, N., & Smith, K. (2016). Eliminating unpredictable linguistic variation through interaction (in preparation).

Fischer, J. L. (1958). Social influences on the choice of a linguistic variant. *Word, 14*, 47–56.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39*, 218–245.

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

Garrod, S., & Pickering, M. J. (2013). Dialogue: Interactive alignment and its implications for language learning and language change. In P. Binder & K. Smith (Eds.), *The language phenomenon: Human communication from milliseconds to millennia* (pp. 47–64). Berlin: Springer.

Garrod, S. C., & Clark, A. (1993). The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes, 8*, 101–126.

Geffen, S., & Mintz, T. H. (2014). Can you believe it?! Twelve-month-olds use word order to distinguish between declaratives and polar interrogatives. *Language Learning and Development, 11*, 270–284.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language, 32*, 249–268.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge: Cambridge University Press.

Giles, H., & Powesland, P. (1975). *Speech style and social evaluation*. London: Academic Press.

Giles, H., Taylor, D. M., & Bourhis, R. Y. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society, 2*, 177–192.

Givón, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.). *The crosslinguistic study of language acquisition* (Vol. 2, pp. 1005–1028). Hillsdale, NJ: Lawrence Erlbaum.

Goldberg, A. E. (2013). Substantive learning bias or an effect of familiarity? Comment on Culbertson, Smolensky, and Legendre (2012). *Cognition, 127*(3), 420–426.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109–135.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178–186.

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research, 34*, 365–399.

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition, 75*, 27–39.

Heine, B. (1997). *Cognitive foundations of grammar*. Oxford: Oxford University Press.

Hockett, C. F. (1960). The origin of speech. *Scientific American, 203*, 88–96.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: Linguistic applications* (pp. 11–42). Washington, DC: Georgetown University Press.

Hudson Kam, C. L. (2009). More than words: Adults learn probabilities over categories and relationships between them. *Language Learning and Development, 5*, 115–145.

Hudson Kam, C. L. (2015). The impact of conditioning variables on the acquisition of variation in adult and child learners. *Language, 91*, 906–937.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151–195.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59*, 30–66.

Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language, 50*, 182–195.

Kerr, D., & Smith, K. (2016). The spontaneous emergence of linguistic diversity in an artificial language. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 11th international conference (EVOLANG 11)*. . Online at<http://evolang.org/neworleans/papers/112.html>.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology, 2*, 125–156.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, USA, 105*, 10681–10686.

Kirby, S., Griffiths, T. L., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology, 28*, 108–114.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition, 141*, 87–102.

Levelt, W. J. M., & Kelter, S. (1982). Surface forma nd memory in question answering. *Cognitive Psychology, 106*, 78–106.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word recognition and learning: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*, 202–227.

Messenger, K., Branigan, H. P., McLean, J. F., & Sorace, A. (2012). Is young children's passive syntax semantically constrained? Evidence from syntactic priming. *Journal of Memory and Language, 66*, 568–587.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30*, 678–686.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*, 61–64.

Pardo, J. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustic Society of America, 119*, 2382–2393.

Peirce, J. W. (2007). Psychopy – Psychophysics software in python. *Journal of Neuroscience Methods, 162*, 8–13.

Perfors, A. (2016). Adult regularization of inconsistent input depends on pragmatic factors. *Language Learning and Development, 12*, 138–155.

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language, 81*, 1–15.

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

Pickering, M. J., Branigan, H. P., Cleland, A. A., & Stewart, A. (2000). Activation of syntactic information during language production. *Journal of Psycholinguistic Research, 29*, 205–216.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134*, 427–459.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169–225.

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation, 4*, 203–228.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2009). The role of distributional information in linguistic category formation. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th annual conference of the Cognitive Science Society* (pp. 2564–2569). Austin, TX: Cognitive Science Society.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology, 66*, 30–54.

Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. M. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition, 125*, 49–63.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Samara, A., Smith, K., Brown, H., & Wonnacott, E. (2016). Acquiring variation in an artificial language: Children and adults are sensitive to socially-conditioned linguistic variation (submitted for publication).

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations. *Developmental Science, 6*, 557–567.

Schenkein, J. (1980). A taxonomy for repeating action sequences in natural conversation. In B. Butterworth (Ed.), *Language production* (pp. 21–47). London: Academic Press.

Shuy, R. W., Wolfram, W., & Riley, W. K. (1967). *Linguistic correlates of social stratification in Detroit speech. Tech. rep., U.S. Office of Education final report No. 6-1347*.

Smith, K., Fehér, O., & Ritt, N. (2014). Eliminating unpredictable linguistic variation through interaction. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the Cognitive Science Society* (pp. 1461–1466). Austin, TX: Cognitive Science Society.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition, 116*, 444–449.

Smith, K. H. (1966). Grammatical intrusions in the recall of structured letter pairs: Mediated transfer or position learning? *Journal of Experimental Psychology, 72*, 580–588.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*, 1558–1568.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209–253.

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition, 4*, 357–380.

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. F. (2014). Socially-mediated syntactic alignment. *Language Variation and Change, 26*, 387–420.

Wedel, A. (2007). Feedback and regularity in the lexicon. *Phonology, 24*, 147–185.

Weiner, E. J., & Labov, W. (1983). Constraints on the agentless passive. *Journal of Linguistics, 19*, 29–58.

Winters, J., Kirby, S., & Smith, K. (2015). Linguistic systems adapt to their contextual niche. *Language and Cognition, 7*, 415–449.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language, 65*, 1–14.

Wonnacott, E., Boyd, J. K., Thomson, J., & Goldberg, A. E. (2012). Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Language, 66*, 458–478.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56*, 165–209.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*(5), 414–420.