



Cognitive Science 44 (2020) e12876

© 2020 The Authors. Cognitive Science published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS). All rights reserved.

ISSN: 1551-6709 online

DOI: 10.1111/cogs.12876

The Role of Social Network Structure in the Emergence of Linguistic Structure

Limor Raviv,^a Antje Meyer,^{a,b} Shiri Lev-Ari^{a,c}

^aMax Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

^bRadboud University, Nijmegen, The Netherlands

^cRoyal Holloway University of London, Egham Hill, UK

Received 9 February 2020; received in revised form 12 June 2020; accepted 26 June 2020

Abstract

Social network structure has been argued to shape the structure of languages, as well as affect the spread of innovations and the formation of conventions in the community. Specifically, theoretical and computational models of language change predict that sparsely connected communities develop more systematic languages, while tightly knit communities can maintain high levels of linguistic complexity and variability. However, the role of social network structure in the cultural evolution of languages has never been tested experimentally. Here, we present results from a behavioral group communication study, in which we examined the formation of new languages created in the lab by micro-societies that varied in their network structure. We contrasted three types of social networks: fully connected, small-world, and scale-free. We examined the artificial languages created by these different networks with respect to their linguistic structure, communicative success, stability, and convergence. Results did not reveal any effect of network structure for any measure, with all languages becoming similarly more systematic, more accurate, more stable, and more shared over time. At the same time, small-world networks showed the greatest variation in their convergence, stabilization, and emerging structure patterns, indicating that network structure can influence the community's susceptibility to random linguistic changes (i.e., drift).

Keywords: Language evolution; Linguistic diversity; Grammatical structure; Social structure; Network structure; Input variability

Correspondence should be sent to Limor Raviv, Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525 XD, The Netherlands. E-mail: limor.raviv@mail.huji.ac.il

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Why are languages so different from each other? One possible explanation is that selective pressures associated with social dynamics and language use can influence the emergence and distribution of different linguistic properties—making language typology a mirror of the social environment (Lupyan & Dale, 2016). According to this hypothesis, often referred to as the *Linguistic Niche Hypothesis*, the structure of languages is shaped by the structure of the community in which they evolved. Research in the past decades supports this theory by showing that different types of languages tend to develop in different types of societies (Bentz & Winter, 2013; Lupyan & Dale, 2010; Meir, Israel, Sandler, Padden, & Aronoff, 2012; Nettle, 1999, 2012; Raviv, Meyer, & Lev-Ari, 2019b; Real, Chater, & Christiansen, 2018).

1.1. *Esoteric versus exoteric languages*

Theoretical models of language change typically draw a distinction between two types of social environments—*esoteric* communities and *exoteric* communities—and argue that there are substantial differences in the grammatical structure and overall uniformity of the languages used in such environments (Meir et al., 2012; Milroy & Milroy, 1985; Roberts & Winters, 2012; Trudgill, 1992, 2002, 2009; Wray & Grace, 2007). Specifically, esoteric communities are generally small and tightly knit societies with little contact with outsiders, and therefore have few if any non-native speakers. In contrast, exoteric communities tend to be much bigger and sparser societies, in which there is a higher degree of language contact and more interaction with strangers and, consequently, also a higher proportion of non-native speakers.

Importantly, computational models, typological studies, and empirical work on the formation of new sign languages all suggest that esoteric and exoteric settings promote the emergence of different linguistic structures. For example, languages spoken in esoteric environments are claimed to be morphologically more complex, and have higher chances of developing rich and non-transparent systems of case marking and grammatical categories (Lupyan & Dale, 2010). Exoteric languages tend to have fewer and less elaborate morphological paradigms, and they are more likely to express various grammatical relations (e.g., negation, future tense) by using lexical forms (individual words) rather than inflections (affixes). That is, there seems to be a greater pressure for creating simpler and more systematic languages in exoteric compared to esoteric settings (Nettle, 2012; Trudgill, 2009; Wray & Grace, 2007). This is presumably because (a) members of exoteric communities are more likely to interact with strangers, resulting in a communicative pressure in favor of generalization and transparency, and (b) there is a relatively high proportion of adult second language learners in exoteric communities, who often struggle with learning complex and opaque morphologies.

Exoteric and esoteric languages are also claimed to show different rates of convergence. Members of esoteric communities are highly familiar with each other and share much common ground, which often entails more alignment and stronger conservation of

existing linguistic norms (Milroy & Milroy, 1985; Trudgill, 2002). Yet this high degree of familiarity between members of esoteric communities can preserve variation and reduce the pressure to establish *new* norms in the early stages of language development, as was found in the case of emerging sign languages (Meir et al., 2012). Specifically, new sign languages that develop in esoteric contexts tend to exhibit more variability across speakers, more irregularities, and overall greater context dependence in comparison to sign languages developed in exoteric contexts. In other words, because members of exoteric communities are far less connected to one another and typically share less common ground with each other, such settings can increase the need for conventions and conformity in the early stages of language emergence, but hinder its preservation later on.

1.2. Teasing apart conflating social factors

The distinction between exoteric and esoteric communities relies on several parameters, namely, community size (small vs. big), network structure (highly connected vs. sparsely connected), and the proportion of adult non-native speakers in the community (low vs. high). These social parameters are naturally confounded in real-world environments (e.g., smaller groups also tend to be highly connected), making it hard to evaluate the unique contribution of each of these factors to the observed pattern of results (i.e., that languages used in exoteric contexts have simpler and more systematic morphologies; Lupyan & Dale, 2010). That is, we currently know very little about how each of these properties affects the structure of languages independently, and whether all features are equally influential in shaping linguistic patterns. Disentangling these social features from one another is important for understanding how exactly languages adapt to fit their social environments, and for assessing the individual role of each factor.

Several computational models have attempted to isolate a specific parameter associated with the difference between esoteric and exoteric communities, and to manipulate it separately from the others in order to examine its effects on various linguistic outcomes (Dale & Lupyan, 2012; Fagyal, Swarup, Escobar, Gasser, & Lakkaraju, 2010; Gong, Baronchelli, Puglisi, & Loreto, 2012; Lou-Magnuson & Onnis, 2018; Spike, 2017; Vogt, 2007, 2009; Wichmann, Stauffer, Schulze, & Holman, 2008). Such models generally suggest that different properties of esoteric and exoteric societies are associated with different pressures, yet often report conflicting results due to differences in model setup and parameter selection. For example, similar computational simulations examining the effect of community size can yield opposite results if agents' learning strategies are defined differently (Wichmann et al., 2008); when agents are assumed to copy globally (i.e., from all other agents in their network), larger groups seem to show slower rates of language change, yet when agents are assumed to copy more locally (i.e., from their closest neighbors), community size has no effect. Therefore, while computational models are valuable for teasing apart different social features, they should be tested against experimental data.

Recently, a behavioral study focused on the role of community size, one of the features differentiating between esoteric and exoteric communities, and tested its individual effect on language emergence by contrasting languages created in the lab by big and small

communities, while keeping all other social properties equal (Raviv et al., 2019b). Results showed that groups of eight interacting participants created more systematic languages, and did so faster and more consistently than groups of four interacting participants. The languages developed in the larger groups were more structured (i.e., more compositional) compared to those developed in smaller groups. This finding was explained by the fact that larger groups faced a greater communicative challenge (due to more input variability). These results are in line with the cross-linguistic observations and the theoretical models reported above, and suggest that at least some of the linguistic differences between exoteric and esoteric languages can indeed be attributed to differences in community size. As such, the study provided the first experimental evidence that community size has a unique and causal role in shaping linguistic patterns.

1.3. The postulated role of network structure

What about the other social features that differentiate between esoteric and exoteric communities? Does network structure also have a unique effect, above and beyond community size? An important feature of esoteric societies is their dense nature, in which members are typically connected via strong ties (i.e., family, close friends), and most if not all members of the community are familiar with one another. In contrast, exoteric societies are much sparser, and typically include many weak ties (i.e., acquaintances) and many members that never interact (i.e., strangers). This difference in network connectivity means that members of exoteric societies generally have fewer opportunities to develop common ground and globally align with each other (given that many of them will rarely or never meet), potentially resulting in more variability in the entire network.

Indeed, recent work on the cultural evolution of technology found that an increase in sparse connections from a state of high density (perhaps due to more geographical spread) leads to more innovations and to more diversity in the community (Derex & Boyd, 2016). In this study, well-connected populations were less likely to produce complex technological solutions because of the ability to learn from all members and quickly converge on a local optimum, reducing exploration and cultural diversity in the population. In contrast, individuals in partially connected populations were more likely to progress along different paths of technological accumulation, leading to larger and more diverse technological repertoires and eventually to more complex solutions. These findings complement a long line of work on the prevalence and spread of innovations in social networks, which suggest that sparser ties generally promote more innovations and more variability. Specifically, work on social network structure shows that weak ties in sparser networks provide individuals with access to information, beliefs, and behaviors beyond their own social circle, making the presence and prevalence of weak ties important for cultural innovation, technological accumulation, and the transmission and spread of ideas, behaviors, and norms (Bahmann, 2014; Granovetter, 1983; Liu, Madhavan, & Sudharshan, 2005).

Additionally, weak ties between members of sparser communities can affect the process of conventionalization, as they may entail less language stability, more variability,

and more potential for changes. In contrast, strong ties between members of dense communities can inhibit language change and increase linguistic conformity: tight-knit connections often function as a conservative force, preserving and amplifying existing norms and resisting external pressures to change (Granovetter, 1983; Milroy & Milroy, 1985; Trudgill, 2002, 2009). That is, denser communities may exhibit stricter maintenance of group conventions and therefore more preservation of linguistic norms, even when these norms are relatively complex and irregular (Trudgill, 2002, 2009). However, even though dense networks are postulated to show more stability, once a change does occur it is more likely to quickly spread to the entire community. This is because individuals are more likely to copy the behavior of strong than weak ties (Centola, 2010) and the propagation of variants is typically faster in dense networks than in sparser networks (Centola, 2010; Milroy & Milroy, 1985; Trudgill, 2009). Importantly, sparser networks' difficulty to converge can trigger a stronger need for generalizations and regularizations, which may eventually lead to the creation of more systematic languages (Raviv et al., 2019b; Wray & Grace, 2007).

Although network structure is postulated to have an important effect in shaping linguistic patterns, to date there is no experimental evidence demonstrating its causal role in language complexity. As such, the theoretical claims described above remain hypothetical or anecdotal, and it is still unclear whether and how languages actually change in populations with different types of network structures. The goal of the current study was to fill in this gap in the literature and experimentally test the effect of social network structure on the emergence of new languages using a similar paradigm to that used in Raviv, Meyer, and Lev-Ari (2019a) for demonstrating community size effects.

1.4. Computational evidence for network structure effects in language change

While experimental data are currently lacking, several computational models have examined the effect of social network structure using agent-based simulations. These models typically examine populations of communicating agents in three different types of networks: (a) dense, fully connected networks, in which all agents are connected to each other; (b) small-world networks, which are sparser in comparison to fully connected networks (i.e., there are fewer connections between agents), but in which most "strangers" are indirectly linked by a short chain of shared connections (Watts & Strogatz, 1998); and (c) scale-free networks, which are also characterized by sparsity and short paths but whose distribution of connections follows a power law (i.e., most agents have few connections, yet some agents, the "hubs," have many; Barabási & Albert, 1999).

A typical interaction in such models consists of two agents, who are randomly selected depending on the networks' available connections and their likelihood. Then, one agent (the producer) produces a linguistic variant (e.g., a vowel, word, or phrase) based on their inventory at the time of the interaction, and the other agent (the receiver) updates their own inventory based on that production and whether it is novel or familiar. This simple type of communication and learning (i.e., updating agent's representations) is repeated for many iterations, allowing researchers to observe how variants spread and change over time in a given network. Importantly, the vast majority of these models do not examine the

complexity or the systematicity of communication systems themselves, but rather focus only on the formation of linguistic conventions. This is done either by examining the time it takes for a population of agents to converge on a single linguistic variant or a shared lexicon, or by examining the degree of global alignment in the population after a fixed amount of time.

In most cases, computational models support the claim that differences in the structural properties of networks can lead to differences in convergence rates and in the spread of variants in the population. Specifically, multiple models report that denser networks show more successful diffusion of innovations compared to sparser networks, and that extra-dense networks (e.g., fully connected) typically converge most rapidly (Fagyal et al., 2010; Gong et al., 2012; Ke, Gong, & Wang, 2008). In addition, the existence of “hubs” (i.e., highly connected agents) in scale-free networks was shown to improve convergence and uniformity by advancing the spread of innovations to all agents in the community (Fagyal et al., 2010; Zubek et al., 2017). However, one model suggested that, as long as networks have small-world properties (i.e., as long as “strangers” are indirectly linked by a short chain of shared connections), the network’s specific configuration plays a minor role in the formation of conventions (Spike, 2017).

Interestingly, two models did examine the structure of languages themselves, and they both report that network structure affected linguistic structure in some way. One model looked at the origin and evolution of linguistic categorization of color terms and found that scale-free networks were the fastest to develop color categories, and that those categories were more structured and more efficient compared to those developed in other types of networks (Gong et al., 2012). The second model introduced comprehensive, real-world mechanisms of social learning and language change, and looked at the creation and maintenance of complex morphology (Lou-Magnuson & Onnis, 2018). The results of this model showed that more transitive networks (i.e., with a higher degree of “intimate” connections) were more likely to develop languages with complex morphological structures. Moreover, fully connected networks showed the highest levels of complexity, regardless of community size.

Together, computational models generally support the hypothesis that network structure can affect linguistic outcomes. They show that sparser networks tend to exhibit more structured languages but overall less convergence compared to dense networks, and they suggest that the existence of “hubs” can further promote systemization and alignment. However, such computational models need to be further tested against empirical data obtained from human participants, seeing as they often lack ecological validity in terms of agents’ cognitive capacities (e.g., agents have an unlimited memory capacity) or their behavior (e.g., agents update their inventories after every interaction by overriding all previous variants). As such, the causal role of network structure warrants further experimental validation.

2. The current study

Here, we experimentally tested the individual effect of network structure using a group communication paradigm (Raviv, Meyer, & Lev-Ari, 2019a, 2019b). We examined the

formation of new languages that were developed by different micro-societies that varied in their network structure. Community size was kept constant across conditions, such that all networks were comprised of eight participants, yet differed in their degree of connectivity (i.e., how many people each participant interacted with) and homogeneity (i.e., whether all participants were equally connected). Specifically, we contrasted three different types of networks, which are typically used in computational models (Fig. 1; see Section 3.4 for more details):

1. **Fully connected network (Fig. 1A):** This network is maximally dense, such that all possible connections are realized (i.e., all participants in the group get to interact with each other). It is also homogenous, as every participant has the same number of connections (i.e., seven people). This type of network resembles early human societies, hunter-gatherer communities, and some villages, yet it is overall rare nowadays (Coward, 2010; Johnson & Earle, 2000).
2. **Small-world network (Fig. 1B):** This network is also relatively homogenous such that everyone has approximately the same number of connections (i.e., either three or four other participants), yet it is much sparser than the fully connected network and realizes only half of the possible connections. Importantly, this network has the small-world property of “strangers” being indirectly linked by a short chain of individuals (Watts & Strogatz, 1998). For example, participants G and H never directly interact, but they are indirectly connected via participants B, D, and F, so that innovations can still spread across the group and conventions can be formed.
3. **Scale-free network (Fig. 1C):** This network is equally sparse as the small-world network, and it has the same number of possible connections overall. However, it is not homogenous: not everyone has the same number of connections. While some agents are highly connected, others are more isolated. The distribution of connections in this network roughly follows a power-law distribution (Barabási & Albert, 1999), with few participants having many connections, and a tail of participants having very few connections. For example, participant A is the “hub” and interacts with almost everyone in the group, while participants D and E are more isolated.

Across conditions, the participants’ goal was to communicate successfully with each other using only an artificial language they created during the experiment. Participants in the same group interacted in alternating pairs according to the structural properties of their allocated network condition (see Section 3.4). In each communication round, paired partners took turns in describing novel scenes of moving shapes, such that one participant produced a label to describe a target scene, and their partner guessed which scene they meant from a larger set of scenes (see Section 3.3; Fig. 2).

Over the course of the experiment, we analyzed the emerging languages using several measurements (see Section 3.5): (a) communicative success, reflecting guessing accuracy; (b) convergence, reflecting the degree of global alignment in the network; (c) stability, reflecting the degree of change over time; and (d) linguistic structure, reflecting the degree of compositional label-to-meaning mappings in participants’ languages. These measures are all related to real-world properties of natural languages: Our measure of convergence

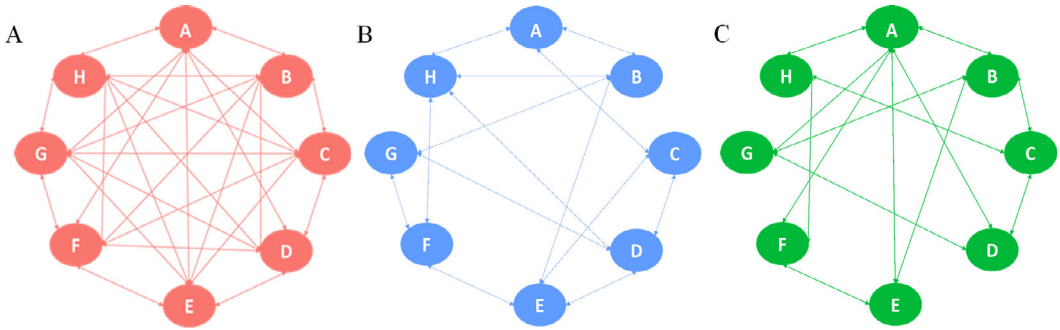


Fig. 1. Network structure conditions. We tested groups of eight participants who were connected to each other in three different setups: a fully connected network (A), a small-world network (B), and a scale-free network (C).

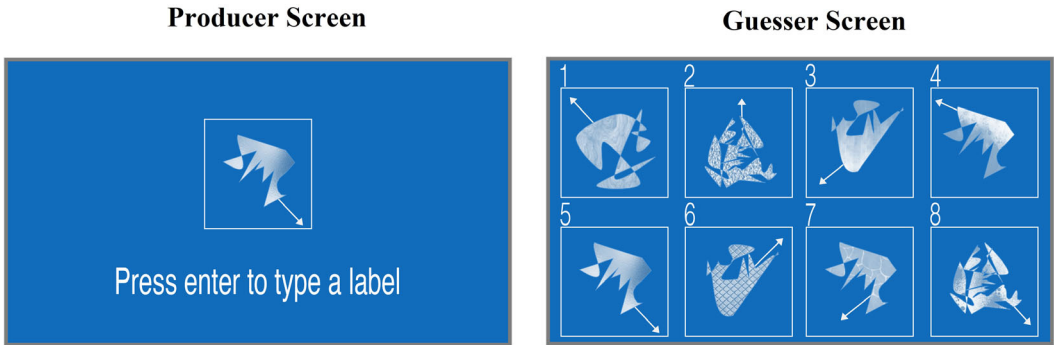


Fig. 2. Example of the computer interfaces in a single interaction during the communication phase. The producer saw the target scene on their screen (A), while the guesser was presented with a grid of eight different scenes on their screen (the target and seven distractors; B). The producer typed a description for the target scene using the artificial language, and the guesser pressed the number associated with the scene they thought their partner was referring to. Paired participants alternated between the roles of producer and guesser. Note that scenes were dynamic events that included a moving shape. The arrows represent the direction of motion.

reflects language uniformity (i.e., the number of dialects in the community and how much people’s lexicons differ from one another); our measure of communicative success is related to mutual understanding; our measure of stability can be taken to reflect languages’ rate of change (i.e., how fast innovations spread in the network); and our measure of linguistic structure can capture various grammatical properties, such as the systematicity of inflectional paradigms and the number of irregulars in a given language. Looking at these four measures enabled us to characterize the emerging languages and to consider how various linguistic properties change over time depending on network structure.

Our predictions with regard to each measure are summarized in Table 1. Our main prediction was that sparser networks would develop more compositional languages, as a

Table 1
Predictions for each measure in the current experiment

Measure	Network Type			Prediction
	Fully Connected (FC)	Small-World (SW)	Scale-Free (SF)	
Input variability		More input variability in sparse networks		FC < SF < SW
Linguistic structure	Sparse networks = more variability → More pressure for generalization and systematicity		“Hubs” can further promote the spread of systematic languages	FC < SW < SF
Convergence	Sparse networks = more variability, more strangers <i>BUT</i> If sparse networks = more systematic languages → Similar levels of convergence		Less convergence	FC = SW = SF
Stability	Dense networks = less diversity → More/faster stability	Sparse networks = more variability → Less/slower stability	more innovations	FC > SW = SF
Communicative success		No difference between conditions		FC = SW = SF

Note The predictions in the table concern the final languages. As described in detail in the text, there could be differences across conditions in the rate of achieving these final outcomes. For example, we predicted that languages in all conditions would eventually show convergence, but we predicted it to occur faster in FC networks.

result of higher levels of input variability and diversity in such networks, which increase the pressure for generalization and systematization (Lou-Magnuson & Onnis, 2018; Raviv et al., 2019b; Wray & Grace, 2007). We also predicted that scale-free networks would show higher compositionality levels compared to small-world networks, since the existence of “hubs” in scale-free networks can further increase the chances of a compositional innovation spreading to the entire population (Fagyal et al., 2010; Gong et al., 2012; Zubek et al., 2017). That is, we predicted that scale-free networks would show the highest degree of linguistic structure (thanks to the “hub”), followed by small-world networks, and then by fully connected networks. We also expected the difference in linguistic structure to be linked to the degree of input variability in dense versus sparse networks: Scale-free and small-world networks should show higher levels of input variability compared to fully connected networks, but the “hub” in scale-free networks may help reduce variability compared to small-world networks by increasing convergence.

Based on the results of Raviv et al. (2019b), we hypothesized that the emergence of more structured languages in sparser networks would promote convergence in such networks (i.e., it should be easier to converge on more systematic variants). That is, while computational models suggest that sparser networks show less convergence in comparison to fully connected networks (given that some participants never interact with each other), we hypothesized that the creation of more structured languages in such networks would facilitate global alignment and lead to similar levels of convergence across networks. Moreover, scale-free networks may exhibit even better global alignment thanks to the existence of a “hub.” In other words, if our prediction about linguistic structure is correct and sparser networks create more systematic languages, then convergence levels should be the same across dense and sparse networks. Otherwise, there should be relatively less convergence in sparser networks.

As for stability, we predicted that sparser networks would be less stable than dense networks, given that there is a higher chance of innovations occurring in sparser networks and more variability overall (Derex & Boyd, 2016), and that changes take longer to stabilize in sparser networks (Ke et al., 2008). As such, we expected to see a difference in the rates of stabilization across conditions, with fully connected networks showing faster stabilization (i.e., less changes over rounds) compared to small-world and scale-free networks. Nevertheless, we expected similar levels of communicative success across all conditions, with all interacting members being equally good at understanding each other.

3. Methods

3.1. Participants

We collected data from 168 adults (mean age = 24.6 years, $SD = 8.1$ years; 132 women), comprising 21 groups of eight members (seven groups in each of the three conditions). Participants were paid 40€ or more depending on the time they spent in the lab (between 270 and 315 min, including a 30-min break). All participants were native Dutch

speakers. Ethical approval was granted by the Faculty of Social Sciences of the Radboud University Nijmegen.

3.2. *Materials*

The materials used in this experiment were identical to those used in our earlier studies (Raviv et al., 2019a, 2019b). Specifically, we created 23 visual scenes that varied along three semantic dimensions: shape, angle of motion, and fill pattern. Each scene included one of four novel unfamiliar shapes, which moved repeatedly in a straight line from the center of the frame in a given direction (i.e., in an angle chosen from a range of possible angles). The shapes were created to be novel and ambiguous in order to prevent easy labeling with existing words. While the dimension of shape included four distinct categories, angle of motion was a continuous feature that could be parsed and categorized by participants in various ways. Additionally, the shape in each scene had a unique blue-hued fill pattern, giving scenes an idiosyncratic feature. Therefore, the meaning space promoted categorization and structure along the dimensions of shape and motion, while still allowing participants to adopt a holistic, unstructured strategy where scenes are individualized according to their fill pattern.

3.3. *Procedure*

The procedure employed in this experiment was identical to that of Raviv et al. (2019b), except for the fact that all groups were comprised of eight participants, and were split up into pairs at the beginning of each communication round depending on their allocated network structure (see Section 3.4; Appendix A). Below we summarize the relevant details.

Participants were told they were about to create a new fantasy language (“Fantasietaal” in Dutch) in the lab and use it in order to communicate with each other about different novel scenes. Participants were not allowed to talk, gesture, point, or communicate in any other explicit way besides the fantasy language and their assigned laptop. Participants’ letter inventory was restricted and included a hyphen, five vowel characters (a, e, i, o, u), and 10 consonant characters (w, t, p, s, f, g, h, k, n, m), which participants could combine freely.

The experiment had 16 rounds, and it included three phases: a group naming phase (round 0), a communication phase (rounds 1–7; rounds 9–15), and a test phase (round 8; round 16).

In the initial naming phase (round 0), participants came up with novel nonsense words to describe eight initial scenes, so that the group had a few shared descriptions to start with. For each of the eight initial scenes, one of the participants was asked to use their creativity and describe it using one or more nonsense words. Participants took turns in describing the scenes, so that the first scene was described by participant A, the second scene was described by participant B, and so on. Importantly, no use of Dutch or any other language was allowed, and participants were instructed to come up with novel nonsense labels. In order to establish mutual knowledge, we presented the scene-description pairings to all participants three times in a random order.

Following the naming phase, participants played a communication game (the communication phase; Fig. 2): The goal was to be communicative and earn as many points as possible as a group, with a point awarded for every successful interaction. The experimenter stressed that this was not a memory game, and that participants were free to use the labels produced during the group naming phase, or create new ones. In each communication round, paired participants interacted with each other 23 times, with participants alternating between the roles of producer and guesser. In a given interaction, the producer saw the target scene on their screen (Fig. 2A) and produced a description for it. Then, they rotated their screen and showed the description (without the target scene) to their partner, the guesser. The guesser was presented with a grid of eight scenes on their screen (the target and seven distractors; Fig. 2B), and had to select the scene they thought their partner referred to. Both participants then received feedback on whether their interaction was successful or not, including the target scene and the selected scene. The number of different target scenes increased gradually over the first six rounds (from eight initial scenes to a total of 23 scenes, with three new scenes introduced at each round), such that participants needed to refer to more and more new scenes as rounds progressed (Raviv et al., 2019a).

At the end of the seventh communication round, participants completed an individual test phase (round 8), in which they were presented with all scenes one by one in a random order, and had to type their descriptions for them using the fantasy language. After the test, participants received a 30-min break and then reconvened to complete seven additional communication rounds (rounds 9–15) and a test round (round 16). At the end of the experiment, all participants filled out a questionnaire about their performance and were debriefed by the experimenter.

3.4. Network properties

We created three different network structures: a fully connected network, a small-world network, and a scale-free network. Each network was comprised of eight individuals (also referred to as nodes or agents), but differed in how these individuals were connected to one another. Fig. 1 shows the configuration of each network. Appendix A includes a detailed description of the order of interactions among pairs in each network condition. These networks can be described using formal measures that are typically used in graph theory. Below we characterize the three networks used in this study in detail and compare them based on the following three measures (see Tables 2 and 3).

Table 2
Comparison of networks' density and global clustering coefficients

	Fully Connected	Small-World	Scale-Free
Number of realized connections	28/28	14/28	14/28
Network density (%)	100	50	50
Global clustering coefficient	1	0.1667	0.4167

Table 3
Comparison of nodes' betweenness centrality across networks

Node	Fully Connected		Small-World		Scale-Free	
	Number of Connections	Betweenness Centrality	Number of Connections	Betweenness Centrality	Number of Connections	Betweenness Centrality
A	7	0	3	0.0476	6	0.3254
B	7	0	4	0.1190	4	0.1111
C	7	0	3	0.0476	3	0.0635
D	7	0	4	0.1190	3	0.0397
E	7	0	4	0.1190	3	0.0238
F	7	0	3	0.0476	3	0.0238
G	7	0	3	0.0476	3	0.0159
H	7	0	4	0.1190	3	0.0635

3.4.1. Network density

This measure reflects the proportion of possible ties which are actualized among the members of a given network. It is measured as the ratio between the number of actual connections in the network and the number of all possible connections (Granovetter, 1976). A possible connection is one that could potentially exist between every two nodes. In a network with n individuals, the number of possible connections is $n*(n - 1)/2$. By contrast, an actual connection is one that really exists in the given network. In a fully connected network where all possible connections are realized, density equals 1 (i.e., 100% connectivity). In a totally isolated network, in which there are no connections between nodes, density equals 0 (i.e., 0% connectivity). All other networks have density values between 0 and 1 (e.g., 0.5, or 50% connectivity, in our experiment).

3.4.2. Global clustering coefficient

This measure, also referred to as transitivity, reflects the degree to which nodes in the network tend to cluster together. In social networks, this measure indicates whether an individual's connections also tend to be connected to each other. In other words, it is the probability that two of one's friends are friends themselves. The global clustering coefficient equals 1 in a fully connected network where everyone knows everyone else, but has typical values in the range of 0.1–0.5 in many real-world networks (Girvan & Newman, 2002). For a given network, this measure is calculated in the following way: For a given node i , the local clustering coefficient is the ratio between the number of realized connections in the neighborhood of node i and the number of all possible connections in that neighborhood if it was fully connected. The average of all nodes' local clustering coefficients yields the global clustering coefficient of the entire network (Watts & Strogatz, 1998).

3.4.3. Betweenness centrality

This measure reflects a node's centrality, that is, how necessary a specific node is for the communication between all the other nodes in the network. In social networks, this measure identifies the most important or influential individuals in the network. That is,

having a high betweenness centrality value suggests that the node is necessary for mediating connections between otherwise unconnected nodes. It is calculated in the following way: For a given node i , betweenness centrality is the number of times node i acts as a bridge along the shortest path between two other nodes (i.e., the number of shortest paths that pass through node i).

3.4.4. Condition 1: Fully connected network

In this condition, depicted in Fig. 1A, all individuals in the network get to interact with one another. As such, all possible connections in the network are realized, and the network is maximally dense and maximally clustered (i.e., density and the clustering coefficient both equal 1). Since all individuals are directly connected to all others, the number of connections per node is identical (i.e., seven), and the betweenness centrality of each node equals 0—no individual is necessary for the others to interact. In our experimental paradigm, it takes seven communication rounds for all pairs in the network to interact (see also Appendix A).

3.4.5. Condition 2: Small-world network

In this condition, depicted in Fig. 1B, only half of the possible connections are realized. As such, this network is much sparser than the fully connected one, and its density is only 0.5 or 50%. In addition, all nodes in the network have a similar number of connections, with each individual being connected to either three or four other individuals. An important feature of small-world networks, which is crucially present in our chosen network, is that the neighbors of any given node are also likely to be neighbors of each other (Watts & Strogatz, 1998). Therefore, unconnected nodes (“strangers”) are still linked by short chains of shared acquaintances. Indeed, every pair of individuals in our small-world network is linked by just one other individual, and typically there is more than one possible mediating individual (resulting in fairly similar and relatively low betweenness centrality values for all nodes, i.e., 0.047 and 0.119). For example, while participants G and H are not connected directly, they are nonetheless indirectly connected via participants F, D, and B. In our experimental paradigm, it takes four communication rounds for all pairs in the network to interact (see also Appendix A).

3.4.6. Condition 3: Scale-free network

In this condition, depicted in Fig. 1C, only half of the possible connections are realized, such that the network’s density is identical to that of the small-world network in condition 2 (i.e., 50% connectivity). Scale-free networks are characterized by the same properties as small-world networks, with an additional important property: The distribution of node degree (i.e., the number of connections the node has to other nodes) follows a power-law (Barabási & Albert, 1999). That is, there are many low-degree nodes (individuals with fewer connections) and a few high-degree nodes (individuals with many connections). The less-connected individuals are often indirectly connected via the highly connected agents, who are often referred to as “hubs.” In our selected network, most participants (i.e., six out of eight) have only three connections, one participant has four

connections, and one participant (“A,” the hub) is connected to almost everyone else in the group. Accordingly, this participant has a very high betweenness centrality score compared to all other participants (i.e., 0.32 vs. 0.11, 0.06, 0.03, 0.02, and 0.01), indicating that they are central for the network’s connectivity, and are necessary for connecting the other participants. In our experimental paradigm, it takes six communication rounds for all pairs in the network to interact (see also Appendix A).

3.5. Measures

3.5.1. Communicative success

Communicative success was measured as the binary response accuracy in a given interaction during the communication phase, reflecting comprehension.

3.5.2. Convergence

Convergence was measured as the similarities between all the labels produced by participants in the same group for the same scene in a given round: For each scene in round n , convergence was calculated by averaging over the normalized Levenshtein distances between all labels produced for that scene in that round (Levenshtein, 1966). The normalized Levenshtein distance between two strings is the minimal number of insertions, substitutions, and deletions of a single character that is required for turning one string into the other, divided by the number of characters in the longer string. This distance was subtracted from 1 to represent string similarity, reflecting the degree of shared lexicon and alignment in the group.

3.5.3. Stability

Stability was measured as the similarities between the labels created by participants for the same scenes on two consecutive rounds: For each scene in round n , stability was calculated by averaging over the normalized Levenshtein distances between all labels produced for that scene in round n and round $n + 1$. This distance was subtracted from 1 to represent string similarity, reflecting the degree of consistency in the groups’ languages.

3.5.4. Linguistic structure

Linguistic structure was measured as the correlation between all pair-wise string distances and semantic distances in each participant’s language in a given round. This correlation reflects the degree to which similar meanings are expressed using similar strings (Kirby, Cornish, & Smith, 2008; Kirby, Tamariz, Cornish, & Smith, 2015). First, scenes had a semantic difference score of 1 if they differed in shape, and 0 otherwise. Second, we calculated the absolute difference between scenes’ angles, and divided it by the maximal distance between angles (180 degrees) to yield a continuous normalized score between 0 and 1. Then, the difference scores for shape and angle were added, yielding a range of semantic distances between 0.18 and 2. Finally, the labels’ string distances were calculated using the normalized Levenshtein distances between all possible pairs of labels

produced by participant p for all scenes in round n . For each participant, the two sets of pair-wise distances (i.e., string distances and meaning distances) were correlated using the Pearson product-moment correlation, yielding a measure of systematic structure (Raviv et al., 2019a, 2019b).

3.5.5. Input variability

Input variability was measured as the minimal sum of differences between all the labels produced for the same scene in a given round (Raviv et al., 2019b). For each scene in round n , we made a list of all label variants for that scene. For each label variant, we summed over the normalized Levenshtein distances between that variant and all other variants in the list. We then selected the variant that was associated with the lowest sum of differences (i.e., the “typical” label) and used that sum as the input variability score for that scene, capturing the number of different variants and their relative difference from each other. Finally, we averaged over the input variability scores of different scenes to yield the mean variability in that round.

3.6. Analyses

We used mixed-effects regression models to test the effect of network condition on all measures using the lme4 package (Bates, 2016) in R (R Core Team, 2016). The reported p -values were generated using the Kenward–Roger Approximation via the pbkrtest package (Halekoh & Højsgaard, 2014), which gives conservative p -values for models based on small numbers of observations. All models had the maximal random effects structure justified by the data that would converge, and they are included in full in Appendix B. The data and the scripts for generating the models can be found at <https://osf.io/utjsb/>.

We examined communicative success, stability, convergence, and linguistic structure using three types of models: (I) models that predict changes in the dependent variable with respect to time and network condition, (II) models that compare the different networks’ final levels of the dependent variable at the end of the experiment, and (III) models that predict the variance of the dependent variable with respect to time and network condition. In all models, NETWORK CONDITION was a three-level categorical factor that was simple coded (i.e., similar to dummy coding except that the intercepts correspond to the grand mean), with fully connected groups as the reference level. That is, we separately contrasted the small-world networks and the scale-free networks with the fully connected networks. This type of contrast coding reflected the fact that most theoretical models of language change predict a difference between dense networks and sparse networks in general. Accordingly, our main prediction regarding the effect of network structure was that the two sparsely connected networks would differ from the fully connected network, which is most accurately captured in the selected coding scheme. Importantly, none of the results reported below changed when using a different coding scheme with a different reference level (e.g., when one of the sparser networks was used as baseline instead).

Models of type (I) predicted changes in the dependent variable over time as a function of network structure. Models for communicative success included data from

communication rounds only (excluding the two test rounds). In models for communicative success, convergence, and stability, the fixed effects were NETWORK CONDITION, ROUND NUMBER (centered), ITEM CURRENT AGE (centered), and the interaction terms NETWORK CONDITION \times ITEM CURRENT AGE and NETWORK CONDITION \times ROUND NUMBER. ITEM CURRENT AGE codes the number of rounds each scene was presented until that point in time and measures the effect of familiarity with a specific scene on performance. ROUND NUMBER measures the effect of time passed in the experiment and overall language proficiency. The random effects structure of models for communicative success, convergence, and stability included by-scenes and by-groups random intercepts and random slopes for the effect of ROUND NUMBER. As linguistic structure score was calculated for each producer over all scenes in a given round, the model for linguistic structure included fixed effects for NETWORK CONDITION, ROUND NUMBER (quadratic¹, centered), and the interaction term NETWORK CONDITION \times ROUND NUMBER, as well as random intercepts and random slopes for the effect of ROUND NUMBER with respect to different producers nested in different groups.

Models of type (II) compared the mean values of the final languages in the last two relevant rounds of the experiment with respect to NETWORK CONDITION. The models for communicative success, stability, and convergence included random intercepts for different groups, and the model for linguistic structure included random intercepts for different producers nested in different groups.

Models of type (III) predicted changes over time in the variance of each measure (i.e., the degree to which different groups differ from each other) as a function of network structure. For linguistic structure, variance was calculated as the square standard deviation in participants' average structure scores across all groups in a given round. For communicative success, convergence, and stability, variance was calculated as the square standard deviation in the dependent variable on each scene across all groups in a given round. All models included fixed effects for NETWORK CONDITION, ROUND NUMBER (centered), and the interaction between them. Models for communicative success, convergence, and stability also included by-scenes random intercepts and random slopes for the effect of ROUND NUMBER.

Following Raviv et al. (2019b), we also examined changes in input variability as a function of time and network structure. This model included fixed effects for NETWORK CONDITION, ROUND NUMBER (quadratic, centered), and the interaction between them, and by-group random intercepts and random slopes with respect to ROUND NUMBER. Finally, we examined changes in linguistic structure over consecutive rounds as a function of input variability. The dependent variable was the difference in structure scores between rounds n and $n + 1$, the fixed effect was MEAN INPUT VARIABILITY at round n (centered), and there were random intercepts for different producers nested in different groups.

4. Results

Below we report the results for each of the four linguistic measures separately. All analyses are reported in full in Appendix B using numbered models, which we refer to here. Fig. 3 summarizes the average performance of different network conditions over the

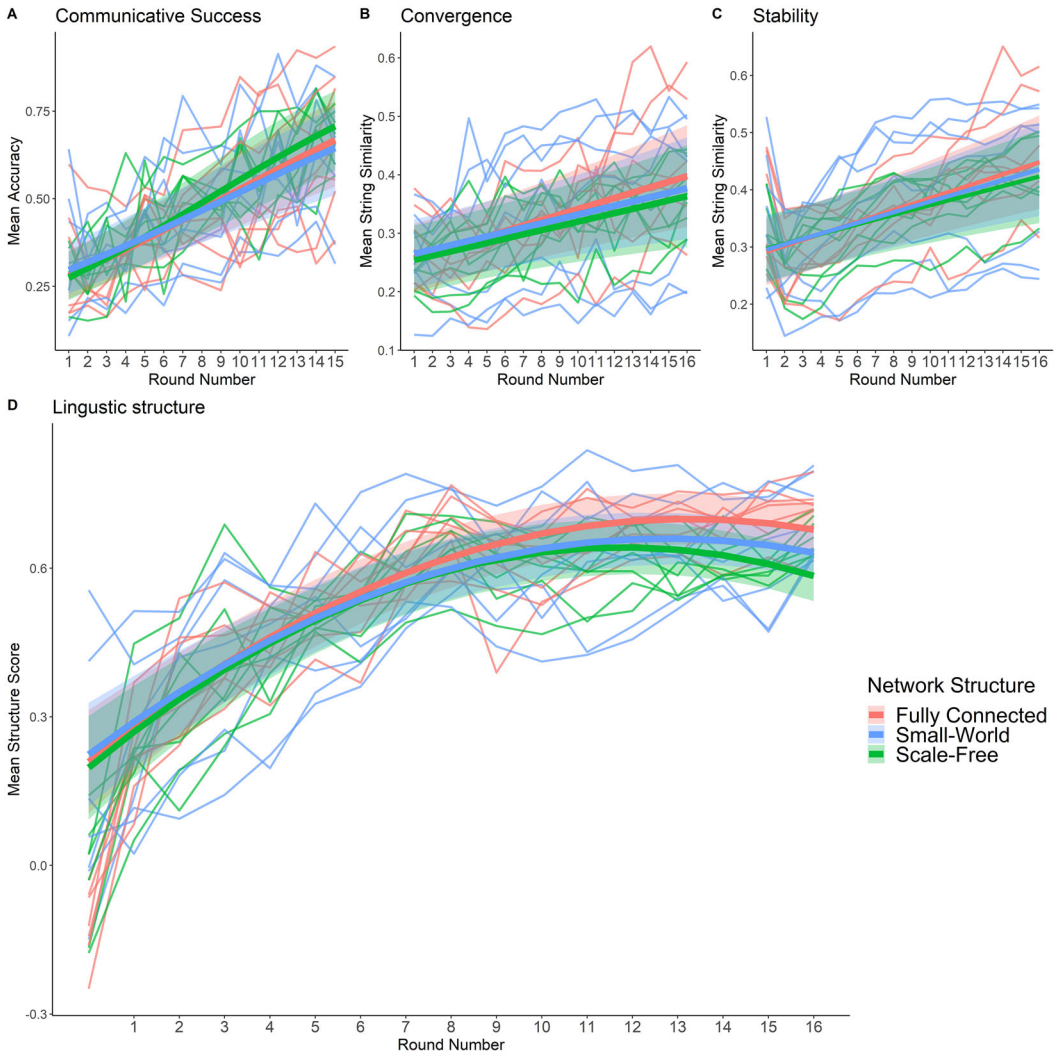


Fig. 3. Changes in (A) communicative success, (B) convergence, (C) stability, and (D) linguistic structure over time as a function of network structure. Thin lines represent average values for each group in a given round. Thick lines represent the models' estimates, and their shadings represent the models' standard errors.

course of the experiment, and Table 4 summarizes the main findings with respect to our predictions.

4.1. Communicative success

Communicative success increased over time (Model 1: $\beta = 0.1$, $SE = 0.01$, $t = 9.74$, $p < .0001$; Fig. 3A), indicating that participants became better at understanding each

other as rounds progressed. All networks shows similar levels of accuracy overall (Model 1: Scale-free vs. fully connected: $\beta = 0.08$, $SE = 0.27$, $t = 0.3$, $p = .76$; Small-world vs. fully connected: $\beta = -0.007$, $SE = 0.27$, $t = -0.03$, $p = .98$), and the increase in accuracy over time was not significantly modulated by network structure (Model 1: Scale-free vs. fully connected: $\beta = 0.01$, $SE = 0.27$, $t = 0.55$, $p = .58$; Small-world vs. fully connected: $\beta = -0.01$, $SE = 0.27$, $t = -0.49$, $p = .62$). Indeed, all networks reached similar levels of accuracy in the final communication rounds (Model 2: Scale-free vs. fully connected: $\beta = 0.26$, $SE = 0.55$, $t = 0.47$, $p = .64$; Small-world vs. fully connected: $\beta = 0.03$, $SE = 0.55$, $t = 0.05$, $p = .96$). No other effect was significant.

As for variance in communicative success, there was no significant difference across network structure conditions (Model 3: Scale-free vs. fully connected: $\beta = 0.001$, $SE = 0.002$, $t = 0.45$, $p = .66$; Small-world vs. fully connected: $\beta = 0.003$, $SE = 0.002$, $t = 1.06$, $p = .3$). Variance in accuracy generally increased over rounds (Model 3: $\beta = 0.001$, $SE = 0.0004$, $t = 2.97$, $p = .006$), but not in scale-free networks (Model 3: $\beta = -0.001$, $SE = 0.0006$, $t = -2.4$, $p = .02$). Together, these results indicate that while groups differed from each other in their accuracy more and more as the experiment progressed (and especially those in the fully connected condition), the difference across groups in the scale-free condition did not change over the course of the experiment.

4.2. Convergence

Convergence increased significantly over rounds (Model 4: $\beta = 0.008$, $SE = 0.001$, $t = 5.42$, $p < .0001$; Fig. 3B), with participants aligning, that is, using more similar labels over time. All networks show similar levels of convergence overall (Model 4: Scale-free vs. fully connected: $\beta = -0.018$, $SE = 0.05$, $t = -0.4$, $p = .7$; Small-world vs. fully connected: $\beta = -0.006$, $SE = 0.05$, $t = -0.14$, $p = .89$), and the increase in convergence over time was not significantly modulated by network structure (Model 4: scale-free vs. fully connected: $\beta = -0.002$, $SE = 0.003$, $t = -0.61$, $p = .55$; small-world vs. fully connected: $\beta = -0.002$, $SE = 0.003$, $t = -0.55$, $p = .58$). Indeed, all networks reached similar levels of convergence by the end of the experiment (Model 5: scale-free vs. fully connected: $\beta = -0.06$, $SE = 0.06$, $t = -1.02$, $p = .32$; Small-world vs. fully connected: $\beta = -0.05$, $SE = 0.06$, $t = -0.92$, $p = .37$). Although there was no significant main effect of item age (Model 4: $\beta = 0.001$, $SE = 0.0008$, $t = 1.43$, $p = .17$), the interaction between network structure and item age was significant, indicating that only fully connected networks showed greater convergence with item age compared to both sparse networks (Model 4: scale-free vs. fully connected: $\beta = -0.004$, $SE = 0.002$, $t = -2.2$, $p = .04$; small-world vs. fully connected: $\beta = -0.005$, $SE = 0.002$, $t = -2.83$, $p = .01$).

Network conditions significantly differed in their degree of variance overall, with scale-free networks showing the lowest variance, and small-world networks showing the highest variance (Model 6: scale-free vs. fully connected: $\beta = -0.007$, $SE = 0.001$, $t = -5.91$, $p < .0001$; small-world vs. fully connected: $\beta = 0.006$, $SE = 0.001$, $t = 5.07$, $p < .0001$). Variance in convergence increased over rounds (Model 6: $\beta = 0.0008$, $SE = 0.0002$, $t = 4.84$, $p < .0001$), but a significant interaction between round and

network indicated that this was not the case for scale-free networks (Model 6: $\beta = -0.001$, $SE = 0.0003$, $t = -4.28$, $p = .0001$). Together, these results suggest that scale-free networks were most consistent in their convergence behavior, while small-world networks were least consistent and varied from each other in their convergence patterns. That is, while some small-world and fully connected networks reached high levels of convergence, others maintained a high level of divergence throughout the experiment, with participants using their own unique labels. In contrast, scale-free networks behaved fairly similar to each other, and reached relatively similar convergence levels. This pattern is also evident in Fig. 3B, where the blue lines corresponding to individual small-world groups show more spread. Thus, throughout the experiment, small-world groups display both very high and very low convergence values.

4.3. Stability

Stability increased significantly over rounds (Model 7: $\beta = 0.009$, $SE = 0.001$, $t = 6.71$, $p < .0001$; Fig. 3C), with participants using labels more consistently as rounds progressed. All networks show similar levels of stability overall (Model 7: Scale-free vs. fully connected: $\beta = -0.008$, $SE = 0.04$, $t = -0.19$, $p = .85$; Small-world vs. fully connected: $\beta = -0.002$, $SE = 0.04$, $t = -0.06$, $p = .96$), and network structure did not modulate the increase in stability over rounds (Model 7: Scale-free vs. fully connected: $\beta = -0.002$, $SE = 0.003$, $t = -0.67$, $p = .51$; Small-world vs. fully connected: $\beta = -0.001$, $SE = 0.003$, $t = -0.36$, $p = .73$). Indeed, all networks reached similar levels of stability by the end of the experiment (Model 8: Scale-free vs. fully connected: $\beta = -0.05$, $SE = 0.06$, $t = -0.82$, $p = .42$; Small-world vs. fully connected: $\beta = -0.05$, $SE = 0.06$, $t = -0.85$, $p = .4$). Although there was no significant main effect of item age (Model 7: $\beta = 0.002$, $SE = 0.0008$, $t = 2.07$, $p = .0516$), the interaction between network structure and item age was significant, indicating that only fully connected networks showed more stability with item age (Model 7: Small-world vs. fully connected: $\beta = -0.005$, $SE = 0.002$, $t = -3.06$, $p = .006$).

Additionally, and as in the case of convergence, network conditions significantly differed in their degree of variance overall, with scale-free networks showing the lowest variance, and small-world networks showing the highest variance (Model 9: Scale-free vs. fully connected: $\beta = -0.006$, $SE = 0.001$, $t = -6.35$, $p < .0001$; Small-world vs. fully connected: $\beta = 0.005$, $SE = 0.001$, $t = 5.65$, $p < .0001$). Even though there was no significant increase in variance in stability over rounds (Model 9: $\beta = 0.003$, $SE = 0.0002$, $t = 1.64$, $p = .11$), a significant interaction between Round number and Network condition indicated that variance increased less over time in scale-free networks (Model 9: $\beta = -0.0006$, $SE = 0.0002$, $t = -2.77$, $p = .009$). In other words, while scale-free networks were most consistent in their behavior, and even more so as the experiment progressed, small-world networks varied most from each other in their stabilization patterns. This pattern is also visually evident in Fig. 3C, where the blue lines corresponding to individual small-world groups show more spread; that is, throughout the experiment, small-world groups display both very high and very low stability values.

4.4. Linguistic structure

Linguistic structure significantly increased over rounds (Model 10: $\beta = 6.39$, $SE = 0.36$, $t = 17.51$, $p < .0001$; Fig 3D), with participants' languages becoming more systematic over time. The increase in structure over time was nonlinear and leveled off in later rounds (Model 10: $\beta = -2.92$, $SE = 0.25$, $t = -11.76$, $p < .0001$). All networks shows similar levels of linguistic structure overall (Model 10: Scale-free vs. fully connected: $\beta = -0.03$, $SE = 0.04$, $t = -0.82$, $p = .42$; Small-world vs. fully connected: $\beta = -0.02$, $SE = 0.04$, $t = -0.48$, $p = .64$), and the increase in structure over time was not significantly modulated by network structure (Model 10: Scale-free vs. fully connected: $\beta = -1.23$, $SE = 0.89$, $t = -1.38$, $p = .18$; Small-world vs. fully connected: $\beta = -0.93$, $SE = 0.89$, $t = -1.04$, $p = .31$). Indeed, all networks reached similar levels of structure by the end of the experiment (Model 11: Scale-free vs. fully connected: $\beta = -0.08$, $SE = 0.04$, $t = -2.05$, $p = .055$; Small-world vs. fully connected: $\beta = -0.05$, $SE = 0.04$, $t = -1.27$, $p = .22$). These findings suggest that networks developed languages with systematic and compositional grammars, and did so to similar extents. To formally test this, we compared the level of structure in the final round of the experiment to chance using the Mantel test with respect to 1,000 random permutations (for a similar procedure, see Kirby et al., 2008). Results indicated that the level of structure in all network conditions was significantly above chance (Fully connected networks: mean structure score = 0.72, mean z -score = 11.45, $p < .0001$; Small-world networks: mean structure score = 0.7, mean z -score = 11.43, $p < .0001$; Scale-free networks: mean structure score = 0.67, mean z -score = 10.98, $p < .0001$). In these systematic languages, participants used complex labels for describing the scenes, with one part typically indicating the shape, and another part typically indicating motion (see Appendix C for multiple examples of final languages created by different groups).

Variance in structure significantly decreased over time (Model 12: $\beta = -0.002$, $SE = 0.003$, $t = -6.13$, $p < .0001$). Additionally, small-world networks were significantly more varied overall in terms of how structured their languages were (Model 12: Small-world vs. fully connected: $\beta = 0.02$, $SE = 0.003$, $t = 5.29$, $p < .0001$). This pattern is also visually evident in Fig. 3D, where the blue lines corresponding to individual small-world groups show more spread; that is, throughout the experiment, small-world groups display both very high and very low structure scores. Given their greater variance to begin with, small-world networks also showed a faster decrease in variance over rounds (Model 12: Small-world vs. fully connected: $\beta = -0.002$, $SE = 0.0007$, $t = -2.86$, $p = .006$). These results suggest that even though small-world networks initially varied most in their level of structure, by the end of the experiment, all networks showed similar and relatively little variability in their level of structure.

Following Raviv et al. (2019b), we also quantified the degree of input variability in each network at a given time point by measuring the differences in the variants produced for different scenes in different rounds. First, we tested whether input variability predicted changes in linguistic structure over consecutive rounds. Our results were in line with the findings of Raviv et al. (2019b) and confirmed that more input variability at round n

induced a greater increase in structure at the following round (Model 13: $\beta = 0.02$, $SE = 0.003$, $t = 6.2$, $p < .0001$). We also found that input variability significantly decreased with time (Model 14: $\beta = -28.15$, $SE = 1.82$, $t = -15.5$, $p < .0001$), but the rate of decrease slowed down in later rounds (Model 14: $\beta = 26.31$, $SE = 1.72$, $t = 15.3$, $p < .0001$). There was also a significant interaction between the linear term of Round number and Network condition (Model 14: Scale-free vs. fully connected: $\beta = 5.85$, $SE = 2.57$, $t = 2.27$, $p = .028$; Small-world vs. fully connected: $\beta = 7.54$, $SE = 2.57$, $t = 2.93$, $p = .005$), showing that input variability decreased more slowly in small-world and scale-free networks than in fully connected networks. Importantly, there was no significant main effect of Network condition (Model 14: Scale-free vs. fully connected: $\beta = 0.07$, $SE = 0.18$, $t = 0.37$, $p = .71$; Small-world vs. fully connected: $\beta = 0.05$, $SE = 0.18$, $t = 0.26$, $p = .8$). This result suggests that, in contrast to our prediction (i.e., that sparse networks would show more variability), there was no effect of network structure on input variability, such that all networks had similar levels of input variability overall. Given the assumed causal relationship between the amount of input variability and the creation of more linguistic structure, the lack of difference in the degree of input variability across the different network conditions may explain why there was no effect of network structure on linguistic structure, as we further discuss below.

5. Discussion

The current study experimentally tested the effect of social network structure on the formation of new languages using a group communication paradigm. We compared the behaviors of groups that varied in their network architecture, contrasting three types of networks: (a) fully connected networks, in which all members interact with each other; (b) small-world networks, which are much sparser and have many members that never interact, although these “strangers” are nevertheless linked indirectly via a short chain of shared connections; and (c) scale-free networks, which are as sparse as small-world networks, but whose members’ distribution of connectivity roughly follows a power law such that one of the participants is highly connected to almost everyone in the network (a “hub”) and others are much less connected.

Based on theoretical and computational models, we generated several predictions (Table 1). First, we predicted that there would be more input variability in sparser networks, given that in such networks, some of the community members never directly interact (i.e., there are more strangers). We hypothesized that this greater input variability and difficulty in convergence would induce a stronger pressure for generalization and systemization, which would result in the sparser networks creating more systematic languages compared to fully connected networks. We further predicted that the emergence of more structured languages in sparser networks would facilitate convergence, allowing members of sparser networks to align on a shared language more easily and therefore resulting in similar convergence to that of fully connected networks. Moreover, we predicted that scale-free networks would develop even more structured languages thanks to the

existence of the hub, who can potentially promote the spread of conventions and systematic innovations. Furthermore, we predicted that sparser networks would stabilize to a lesser extent or more slowly compared to fully connected networks, given that changes take longer to stabilize in sparser networks. Finally, we predicted that all networks would reach similar levels of communicative success, such that across conditions, members that interacted with each other would understand each other equally well.

Table 4 summarizes our experimental results and compares them to our research questions and predictions as presented in Table 1 in the beginning of the paper, and Table 5 summarizes additional results that were obtained in the study but not directly predicted. We found that over time, all groups developed languages that were highly systematic, communicatively efficient, stable, and shared across members. However, there were no significant differences between the three network conditions on any measure with respect to our original predictions (Table 4): All networks showed the same behavioral patterns, had similar degrees of input variability, and reached similar levels of linguistic structure, stability, convergence, and communicative success. While the results for communicative success and convergence are in line with our predictions (i.e., that all networks would show similar levels of communicative accuracy and global alignment), the remaining predictions were not borne out. Below we discuss potential reasons for this.

Although we did not find any significant differences between the three network conditions that were directly relevant to our original predictions (Table 4), we did find several other differences in the networks' patterns (Table 5). First, fully connected networks differed from the two sparser networks (i.e., small-world and scale-free networks) when looking at items' age, a variable that represents familiarity with specific items. Specifically, we found that only fully connected networks showed better convergence and better stability for older items compared to more recently introduced items. Specifically, participants in fully connected networks, but not in the sparser networks, were more aligned on items that were introduced in earlier rounds (and therefore repeated more often), and changed their labels for older items to a lesser extent. This result suggests that network structure might influence patterns of convergence and stability. However, since there was no overall difference between network conditions with respect to Round Number (a variable that represents the overall time passed in the experiment and participants' general experience with the language), these results do not represent strong evidence in favor of a network effect.

One consistent pattern that emerged from all our additional analyses was that small-world networks showed the most variance in their observed behaviors (Table 5), with different small-world networks behaving very differently from one another (not to be confused with the similar levels of input variability within each network). Fully connected networks and scale-free networks were generally similar to other fully connected networks and other scale-free networks, respectively, in terms of their convergence, stability, and linguistic structure levels. However, small-world networks showed a great deal of variance, with different groups in the same condition showing very different levels of these three measures (also evident in Fig. 3, which shows a high degree of dispersity for small-world networks). These results suggest that small-world networks may be more sensitive to random events (i.e., drift). Specifically, frequent interactions among small subgroups can preserve random behaviors more easily,

Table 4
Experimental results versus predictions for each measure

Measure	Network Type			Prediction	Experimental Results
	Fully Connected (FC)	Small-World (SW)	Scale-Free (SF)		
Input variability	More input variability in sparse networks			FC < SF < SW	FC = SW = SF
Linguistic structure	Sparse networks = more variability → More pressure for generalization and systematicity	“Hubs” can further promote the spread of systematic languages		FC < SW < SF	FC = SW = SF
Convergence	Sparse networks = more variability, more strangers <i>BUT</i>	→ less convergence		FC = SW = SF	FC = SW = SF
Stability	If sparse networks = more systematic languages → Similar levels of convergence				
	Dense networks = less diversity → More/faster stability	Sparse networks = more variability, more innovations	Less/slower stability	FC > SW = SF	FC = SW = SF
Communicative success	No difference between conditions			FC = SW = SF	FC = SW = SF

Table 5
Additional experimental results

Measure	Analyses	
	Variance in Behavior	Additional Results
Input variability	—	Input variability decreased more slowly in the sparser networks (small-world and scale-free) compared to the fully connected networks
Linguistic structure	Small-world networks were the least consistent (i.e., showed most variance) in how structured their languages were, but less so over time	—
Convergence	Small-world networks were the least consistent (i.e., showed most variance) in their convergence patterns, and scale-free networks were the most consistent	Only in fully connected networks there was more convergence on older items compared to new ones
Stability	Small-world networks were the least consistent (i.e., showed most variance) in their stabilization patterns, and scale-free networks were the most consistent	Only in fully connected networks were labels for older items more stable than labels for newer items
Communicative success	All networks showed overall similar levels of variance in their accuracy	—

resulting in small-world networks being more likely to fixate on local (and possibly costly) strategies instead of converging on optimal solutions (Bahlmann, 2014; Kurvers, Krause, Croft, Wilson, & Wolf, 2014). Our finding that small-world networks show more variance in their linguistic behaviors raises several predictions worth investigating. First, it suggests that changes in community structure across history that required greater geographical spread and reduced contact may have led to greater diversification in linguistic structure. Second, it might suggest that community structure can predict how likely communities are to exhibit common linguistic features compared to more rare ones (e.g., common vs. uncommon word order). Future research should investigate how community structure can influence the likelihood of a given language to follow or violate common trajectories of language change.

As mentioned earlier, the results of the study differed from those we had predicted (Table 4). We predicted that different networks would show similar levels of convergence, but the rationale behind this prediction was not met. We hypothesized that the similar levels of convergence across networks would be the result of sparser networks initially showing greater input variability (hindering convergence in comparison to the fully connected networks), but that this greater variability would eventually lead sparser networks to create more systematic languages, which would in turn help them overcome this disadvantage. That is, our prediction was based on the idea that different networks would reach a sort of equilibrium between their difficulty to converge and their need to converge. Crucially, this was not the case: All networks showed similar levels of input variability and systematic structure. This discrepancy fits our findings of equal convergence

across conditions: Different networks showed the same convergence patterns because their degree of input variability was the same.

While our results are surprising given the literature reviewed in the Introduction, they are in line with the computational model described in Spike (2017), who concluded that network structure plays a relatively small role in the development and maintenance of linguistic complexity and linguistic norms. This model simulated the process of conventionalization in populations of agents that varied in their community size, network structure, and learning biases (Spike, 2017). While the learning capacity of agents and the size of the population influenced the final outcomes of the model, results from multiple simulations showed that network structure had no apparent long-term effects on language change. Spike (2017) concluded that as long as populations exhibit a small-world property, that is, that the average path length between any two people is small (which is the case in all our three network conditions), the diffusion of variants across the network is sufficient to ensure similar linguistic trends. As in our experimental manipulation, real-world social networks are small-world in nature (Watts & Strogatz, 1998). That is, it is possible that network structure has little to no effect on the formation linguistic trends, at least in relatively natural networks.

However, we believe this interpretation is unlikely to be correct given the theoretical and computational models that argue in favor of network structure effects (Fagyal et al., 2010; Gong et al., 2012; Ke et al., 2008; Lou-Magnuson & Onnis, 2018). We believe it is more likely that the current study did not sufficiently capture the potential role of network structure. One possibility is that network structure interacts with group size in complex ways (as suggested by Lou-Magnuson & Onnis, 2018), and/or that network structure effects only manifest themselves once a certain group size threshold has been crossed. That is, it is possible that our eight-person networks were simply too small, and that running this experiment with bigger networks (e.g., of 200 people) would yield different results. Disentangling the relation between group size and network structure experimentally would require further investigation, potentially using online adaptations of this paradigm, which would allow testing much larger groups of interacting participants.

Another possibility is that, regardless of group size, our network structure manipulation was not strong enough to create meaningful differences between network types, or was not representative of real-world differences between dense and sparse networks. For example, the sparse networks might not have been sparse enough, or the difference between the small-world and scale-free networks might have been too subtle. Notably, the nature of our experimental procedure restricted the specific architecture of sparser networks to a great extent. At any given communication round, each network had to be divided into pairs who played the game simultaneously, with no participant left out. Given this constraint, our choice of possible connections between group members was highly limited: Many possible network configurations did not adhere to this constraint and were therefore inappropriate for our design. For the sake of illustration, imagine designing a four-person network that is sparsely connected, such that only four out of the six possible connections are realized. While there are 15 hypothetical network configurations that meet this definition, only three of them satisfy the condition of being able to be divided into two unique pairs at a given time point

and can therefore be used in our experimental paradigm. In the remaining 12 theoretical networks configurations, one participant would need to be included in two pairs at the same time, or would be without a communication partner. While it is relatively simple to find out which three networks out of the 15 hypothetical four-person networks would be suitable for our design, the problem was exponentially worse with the larger networks used in the current study: For sparser networks with eight individuals and 14 realized connections, there are over 40 million possibilities for network configurations, and only a few of them are suitable for our design. As such, we cannot rule out that the networks selected for this experiment were not representative of real-world sparser networks, and/or had biased characteristics that made them too similar to one another. In other words, it is reasonable to assume that network structure had no effect in the current design because our selected networks did not differ sufficiently from each other. This possibility is supported by the lack of observed differences in input variability across conditions, which stands in sharp contrast with the general consensus that sparser networks should be more diversified (Bahlmann, 2014; Derex & Boyd, 2016; Liu et al., 2005).

The similar levels of input variability across network conditions may, in fact, explain the remaining results of this study. Evidently, the prediction that sparse networks would show more input variability was a key component underlying the predictions for stability and linguistic structure. Since it turned out to be false, it is perhaps not surprising that the predictions that were based on it also turned out to be false. In the case of stability, we hypothesized that more input variability in sparser networks would lead to slower or less stabilization in such networks. Given that there were no differences in input variability between the dense and sparse networks, it is not surprising that they also showed similar degrees of stability over time. In the case of linguistic structure, our prediction for structural differences between network conditions relied on the causal relation between input variability and systematic structure. This relation, that is, that more input variability promotes more linguistic structure, was demonstrated in Raviv et al. (2019b) and confirmed in the current study. We found that, across conditions and experimental rounds, more input variability at time point n induced more structure at time point $n + 1$. Therefore, if sparse networks indeed show greater input variability, they should consequently show more linguistic structure. However, if all networks show similar levels of input variability, they should also show similar levels of linguistic structure—which is what we found in the current study. Together, these results support the idea that network structure had no effect in our study because our selected networks did not differ substantially from each other. Perhaps, a stronger manipulation of network sparsity would have yielded different results. Therefore, more research is required in order to confirm or refute the influence of network structure on linguistic patterns.

We also predicted that scale-free networks would develop even more structured languages than small-world and fully connected networks due to the existence of a highly connected participant (a “hub”), who should potentially promote the spread of systematic variants to the entire community once they emerge (Fagyal et al., 2010; Zubek et al., 2017). This prediction was not met, and scale-free networks showed similar levels of linguistic structure to the other two network types. In retrospect, this discrepancy is very

likely to be the result of the specific properties of our design: Given that all networks in our experiment received the same amount of time for interaction (14 communication rounds in total, see Section 3.3) and given that each communication round included simultaneous communication between pairs, having more connections inevitably resulted in having less time to interact with each connection. Given these features, a highly connected participant would require more rounds to interact with all their possible connections, while a less connected participant would in the meantime repeatedly interact with the same few connections. While such subgroups can be seen as a relevant feature of sparser networks, this configuration also resulted in the highly connected participants interacting less with each of their connections. That is, while the highly connected agent was indeed well connected in the sense that they communicated with almost every person in the group, they were actually *less connected* to each person in terms of their frequency of interactions: The hub interacted approximately twice with each of their connections by the end of the experiment, while the less connected participants interacted among themselves for approximately six times in the meanwhile. From the perspective of the less connected participants, who repeatedly conversed with the same people and only rarely interacted with the hub, the hub could have effectively be seen as “an outsider,” that is, a person they rarely interacted with, and consequently a person who mattered less. That is, our design may have maintained the structural property of the hub but stripped it of their commonly associated social meaning, namely, having greater rather than lesser social importance. If true, it would again suggest that a different design or a different network selection would have revealed different results.

One way of dealing with the methodological issues described above is to move away from our current design and introduce more flexible communication conditions, while maintaining equal experience across all individuals in the group. For example, it is possible to include individual rounds or semi-communicative rounds, in which a participant is not assigned a partner, but nevertheless engages in some form of communicative behavior, for example, with a computer-simulated agent. Alternatively, it is possible to introduce multi-player rounds, in which three participants are assigned to communicate together so that one participant produces a word and the other two participants guess the corresponding scene separately. Such modifications would dramatically improve the flexibility of our paradigm and expand the pool of suitable networks, while also introducing more varied conversational settings. Nevertheless, they introduce new challenges: The degree of input variability (and consequently, the difficulty of convergence) may be reduced if participants can interact with several group members at the same time, and it is not clear how to simulate a computerized participant in a way that mimics human participants’ behavior and produces the same communicative challenges faced by people interacting with a real participants.

Finally, it is worth mentioning that our network structures were fixed and did not change over time. Therefore, our sparse networks differed from real-world sparse networks in the sense that pairs of participants who were not directly connected to each other would in fact *never* interact, and probably figured this out. Some researchers have argued that an important feature of real-world sparse communities is the increased possibility of interacting with strangers, and they treat interaction with strangers as a crucial

mechanism driving morphological simplification (Wray & Grace, 2007). The idea behind this argument is that increasing the chances of interacting with unfamiliar people (with whom you have no shared history) introduces a stronger pressure for creating languages with simpler, transparent, and regular structure (Granito, Tehrani, Kendal, & Scott-Phillips, 2019). In other words, the *potential* of encountering a new member of one's community may be relevant for explaining cross-linguistic differences. One way of testing this hypothesis is by introducing a more dynamic, open-ended network design, for example, by assigning an unexpected connection every few rounds, so that individuals who are not directly connected may nevertheless encounter each other randomly from time to time.

6. Conclusions

The current study attempted to experimentally test the influence of social network structure on emerging languages using a group communication paradigm. We found no effect of network structure on any measure, with fully connected, small-world, and scale-free networks all showing similar patterns of communicative success, convergence, stability, and linguistic structure. We argue that these findings could be traced back to the lack of differences in input variability between network conditions in our design, and that further research is needed in order to confirm or refute the role of network structure on language evolution and language change. Nevertheless, our results show that network structure can significantly affect communities' susceptibility to drift, with small-world networks being more likely to vary from each other and fixate on local strategies.

Acknowledgments

We wish to thank Caitlin Decuyper for programming the experiment, and Phillip Alday, Mark Atkinson, Mattan S. Ben-Shachar, Rona Feldman, Gary Lupyan, Joe Rodd, William Schueller, and José Segovia Martín for discussions and helpful input. Open access funding enabled and organized by Projekt DEAL

Open Research Badges



This article has earned Open Data and Open Materials Badges. Data and materials are available at <https://osf.io/utjsb/>.

Note

1. We included both the linear and the quadratic terms using the `poly()` function in R to avoid collinearity.

References

- Bahlmann, M. D. (2014). Geographic network diversity: How does it affect exploratory innovation? *Industry and Innovation*, 21(7–8), 633–654.
- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509–512.
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2016). *lme4: Mixed-effects modeling with R*; 2010. Available at <https://cran.r-project.org/web/packages/lme4>
- Bentz, C., & Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, 3(1), 1–27.
- Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, 329(5996), 1194–1197.
- Coward, F. (2010). Small worlds, material culture and ancient Near Eastern social networks. *Social brain, distributed mind*, 449–484. British Academy.
- Dale, R., & Lupyán, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems*, 15(03n04), 1150017.
- Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences of the United States of America*, 113(11), 2982–2987.
- Fagyal, Z., Swarup, S., Escobar, A. M., Gasser, L., & Lakkaraju, K. (2010). Centers and peripheries: Network roles in language change. *Lingua*, 120(8), 2061–2079.
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 7821–7826.
- Gong, T., Baronchelli, A., Puglisi, A., & Loreto, V. (2012). Exploring the roles of complex networks in linguistic categorization. *Artificial Life*, 18(1), 107–121.
- Granito, C., Tehrani, J., Kendal, J., & Scott-Phillips, T. (2019). Style of pictorial representation is shaped by intergroup contact. *Evolutionary Human Sciences*, 1, e8.
- Granovetter, M. (1976). Network sampling: Some first steps. *American Journal of Sociology*, 81(6), 1287–1303.
- Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1(1983), 201–233.
- Halekoh, U., & Højsgaard, S. (2014). A Kenward–Roger approximation and parametric bootstrap methods for tests in linear mixed models – The R package pbrktest. *Journal of Statistical Software*, 59(9), 1–32.
- Johnson, A. W., & Earle, T. K. (Eds.). (2000). *The evolution of human societies: From foraging group to Agrarian state*. Stanford, CA: Stanford University Press.
- Ke, J., Gong, T., & Wang, W. S. (2008). Language change and social networks. *Communications in Computational Physics*, 3(4), 935–949.
- 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 of the United States of America*, 105(31), 10681–10686.
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87–102.
- Kurvers, R. H. J. M., Krause, J., Croft, D. P., Wilson, A. D. M., & Wolf, M. (2014). The evolutionary and ecological consequences of animal social networks: Emerging issues. *Trends in Ecology & Evolution*, 29(6), 326–335.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8), 707–710.
- Liu, B.-S.-C., Madhavan, R., & Sudharshan, D. (2005). DiffuNET: The impact of network structure on diffusion of innovation. *European Journal of Innovation Management*, 8(2), 240–262.
- Lou-Magnuson, M., & Onnis, L. (2018). Social network limits language complexity. *Cognitive Science*, 42(8), 2790–2817.

- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, 5(1), e8559.
- Lupyan, G., & Dale, R. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends in Cognitive Sciences*, 20(9), 649–660.
- Meir, I., Israel, A., Sandler, W., Padden, C. A., & Aronoff, M. (2012). The influence of community on language structure: Evidence from two young sign languages. *Linguistic Variation*, 12(2), 247–291.
- Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics*, 21(02), 339–384.
- Nettle, D. (1999). Using social impact theory to simulate language change. *Lingua*, 108(2), 95–117.
- Nettle, D. (2012). Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1597), 1829–1836.
- R Core Team. (2016). *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <http://www.R-project.org/>
- Raviv, L., Meyer, A., & Lev-Ari, S. (2019a). Compositional structure can emerge without generational transmission. *Cognition*, 182, 151–164.
- Raviv, L., Meyer, A. S., & Lev-Ari, S. (2019b). Larger communities create more systematic languages. *Proceedings of the Royal Society B: Biological Sciences*, 286(1907), 20191262.
- Real, F., Chater, N., & Christiansen, M. H. (2018). Simpler grammar, larger vocabulary: How population size affects language. *Proceedings of the Royal Society B: Biological Sciences*, 285(1871), 20172586.
- Roberts, S. G., & Winters, J. (2012). Social structure and language structure: The new nomothetic approach. *Psychology of Language and Communication*, 16(2), 89–112.
- Spike, M. (2017). Population size, learning, and innovation determine linguistic complexity. In G. Gunzelmann, A. Howes, T. Tenbrink & E. J. Davelaar (Eds.), *Proceedings of the 39th annual meeting of the Cognitive Science Society* (pp. 1108–1113). Austin, TX: Cognitive Science Society.
- Trudgill, P. (1992). Dialect typology and social structure. In E. H. Jahr (Ed.), *Language contact: Theoretical and empirical studies* (pp. 195–212). Berlin, NY: Mouton de Gruyter.
- Trudgill, P. (2002). Linguistic and social typology. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 707–728). Malden, MA: Blackwell.
- Trudgill, P. (2009). Sociolinguistic typology and complexification. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 98–109). Oxford, UK: Oxford University Press.
- Vogt, P. (2007). Group size effects on the emergence of compositional structures in language. In F. A. E. Costa, L. M. Rocha, E. Costa, I. Harvey, & A. Coutinho (Eds.), *Advances in artificial life* (pp. 405–414). Berlin, Heidelberg: Springer.
- Vogt, P. (2009). Modeling interactions between language evolution and demography. *Human Biology*, 81(2), 237–258.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440–442.
- Wichmann, S., Stauffer, D., Schulze, C., & Holman, E. W. (2008). Do language change rates depend on population size? *Advances in Complex Systems*, 11(03), 357–369.
- Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117(3), 543–578.
- Zubek, J., Denkiewicz, M., Barański, J., Wróblewski, P., Rączaszek-Leonardi, J., & Plewczynski, D. (2017). Social adaptation in multi-agent model of linguistic categorization is affected by network information flow. *PLoS ONE*, 12(8), e0182490.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix A. Network details.

Appendix B. Models.

Appendix C. Examples of structured languages.