Converging evidence: Network structure effects on conventionalization of gestural referring expressions

Russell Richie[1,*], Matthew L. Hall[2,*], , Pyeong Whan Cho[3], & Marie Coppola[4]

[1] University of Pennsylvania, Psychology
[2] University of Massachusetts - Dartmouth, Psychology
[3] University of Michigan, Psychology
[4] University of Connecticut, Psychological Sciences and Linguistics

* Equal contribution
Corresponding Author: Russell Richie

drrichie@sas.upenn.edu

Abstract:

New languages emerge through interactions among people, yet the role of social network structure in language emergence is not clear, despite research from experimental semiotics, observational fieldwork, and computational modeling. To better understand the effects of social network structure on the formation of conventional referring expressions, we use a silent gesture paradigm that combines the methodological control of experimental semiotics and computational simulations with the naturalistic affordances of the human body, physical environment, and interpersonal communication. We elicited gestural referring expressions from hearing participants randomly assigned to either a richly- or sparsely-connected communicative network. Results demonstrated greater conventionalization among participants in the richly-connected condition, although this effect reverses after accounting for between-condition differences in overall number of communicative interactions. These results provide the first experimental demonstration that communicative network structure causally impacts the conventionalization of referring expressions in human participants, using a communicative modality in which human language naturally arises.

Keywords: gesture; sign language; convention; experimental semiotics; Bayesian data analysis

1 Introduction

How do people agree on what to call things? The emergence of conventionalized referring expressions[1] is a fundamental step in the development of any semiotic system- including human language, which is our focus here. As far back as Saussure (1916/1983), or even Plato's *Cratylus*, scholars have theorized about the impact of social interaction over time in shaping the conventionalization of referring expressions. However, this work was largely speculative; only recently have more empirically-grounded methods arisen for studying the impact of social-communicative structure on emergent linguistic structure. At least three general methods are currently prevalent in the literature: experimental semiotics, observational fieldwork on emerging languages, and computational simulation. Studies using these different methods have not always agreed, as described below, nor are they necessarily informative about the processes by which conventions emerge in *human language*. The present manuscript yields new evidence from a silent gesture task that combines that experimental control of traditional studies in experimental semiotics and computational simulations with the naturalistic affordances of having real human beings with real human bodies in real interaction, communicating in a modality in which human language naturally arises (i.e. visual-gestural communication). We argue that this type of evidence is more informative about the actual processes of linguistic conventionalization compared to approaches that abstract farther away from actual embodied communication.

The central question of the present study is whether the conventionalization of referring expressions will be greater in networks where all members of a population interact with one another, or in networks where all communication involves a central "hub", as illustrated schematically in Figure 1. Before describing the design of the present study, we first review the available evidence about the impact of communicative network structure on the conventionalization of referring expressions, and the naturally emerging systems that motivate the inclusion of these two network structures.

1.1 Evidence From Experimental Semiotics.

Experimental semiotics is an emerging field of research in which human participants are asked to solve some sort of challenge that requires them to communicate with each other, but in which already-conventionalized ways of communication (e.g. speaking, signing, writing) are withheld. Instead, participants must make use of whatever options are available to them to accomplish the task at hand; the object of study is typically the process by which participants arrive at a conventionalized system, and/or the nature of system itself. (For useful overviews of experimental semiotics, see Galantucci, Garrod, & Roberts, 2012; Tamariz, 2017). In these studies, artificial constraints are intentionally imposed on participants to prevent them from defaulting to easier solutions, such as using their native language, because use of an already-conventionalized system would not reveal anything about the process of conventionalization.

[1] We use the term *referring expression* rather than *word* in order to reserve the notion of *word* for referring expressions that have also accrued additional syntactic and phonological features such as grammatical category, syllabic structure, etc. We make no attempt to argue that the referring expressions elicited in our experiment have such features.

The features that are observed in the systems created under such conditions are often described in abstract terms: arbitrariness, iconicity, combinatoriality, compositionality, duality of patterning, etc. Results of this nature are valuable, but they are only one piece of the puzzle. For example, experiments in which participants must signal their intentions by moving a mouse on a computer screen may yield systems that exhibit combinatoriality; however, the fact remains that no human language has evolved in a computer-mouse modality. Conversely, systems like drawing, music, and computer code have many of these features and yet never develop other key aspects of linguistic structure despite long histories of interaction and transmission by humans. Consequently, there is still a need to connect the dots between the *availability* of these abstract concepts and their *actual instantiation* in communicative modalities in which natural languages arise (i.e. sign and speech).

Three previous studies in experimental semiotics have reported that communicative network structure impacts the conventionalization of referring expressions. Fay et al. (2008) asked participants to create drawings to represent various concepts (e.g. microwave, Brad Pitt). Participants were either embedded in a communicative network with seven other people (who all eventually interacted with one another), or in a dyad where neither member interacted with any other. They report, but do not quantify, that for any given item (e.g. Brad Pitt), the drawings produced by the members of the interacting network were more conventionalized than those created by the isolated dyads. A subsequent study by the same group (Fay et al., 2010) quantified conventionalization across the different network conditions by asking naïve participants to rate the similarity between various pairs of drawings. After interaction had had time to exert its maximal effect (i.e. the end of the experiment), drawings made by members of the interacting community were rated as more similar to each other (i.e. more conventionalized) than drawings produced by either isolated pairs or by members of non-interacting communities.

These results are consistent with the hypothesis that conventionalization of referring expressions will be greater in networks where all members of a population interact with one another, relative to networks in which all communication involves a central “hub”. However, the networks in these two studies did not technically differ in their communicative *structure*: all participants in both networks interacted with all other members. Instead, the conditions in the Fay et al. studies differed mainly in the *number of members* of each network (8 in interacting communities vs. 2 in isolated dyads). In addition, we cannot assume that findings from pictorial communication will necessarily translate to other modes of communication; indeed, Fay et al. (2008) explicitly state that their study is about the emergence of pictorial communication, which has different properties than language (Fay et al., 2008, p. 3554).

The third study to have measured the impact of communicative network structure on the emergence of referring expressions (Centola and Baronchelli, 2015) is also useful for illustrating why studying naturalistic human interactions is important. Centola and Baronchelli wanted to understand how various types of network structures impacted the emergence of social conventions in general, of which linguistic behavior is one example. The design of their study involved 510 participants, each of whom was randomly assigned to one of three network structures (spatially-embedded, random, or homogenously-mixed). Participants took part over the internet, and interacted in dyads. The task was to generate a proper name in response to a picture of a human face. After making a guess, participants could see what their partner had guessed, but no further interaction between the partners occurred. Instead, each were assigned a different partner for each subsequent round, with an average of 30 rounds per network. The face in the stimulus picture never changed. Results demonstrated that participants assigned to the homogeneously-mixed network were the only ones to reliably achieve network-wide convergence on a name for the face in the stimulus picture, even after 30 rounds.

These results succeed in demonstrating the relative advantage of the homogeneously-mixed network structure over the spatially-embedded and random networks; however, they do not succeed in demonstrating that the emergence of linguistic referring expressions follows any such process. For example, consider what would probably happen if you and a stranger wanted to refer to a particular person, but neither of you knew that person by name. It is highly unlikely that you would each simultaneously suggest a proper name at random, and then repeat the process with a different stranger if this initial attempt is unsuccessful. Instead, a number of other possibilities would be far more likely: you might overtly suggest a label, and your interlocutor could either accept it, modify it, or suggest another expression instead. Although your initial suggestion might be a proper name if the referent is a person (e.g. "Let's call him Bill"), this strategy is likely to fail if you wanted to refer to anything else (e.g. a novel machine that you might observe when touring a factory, or an abstract sculpture). For nonhuman referents (and potentially also for human referents), you might refer to physical appearance, presumed function, or resemblance to other elements that you believe to be in common ground with your interlocutor. You might also produce deictic or iconic gestures to supplement the information in your speech. These same strategies may also be used to refer to something that's not physically present. This is in fact how humans behave when they share a language with their interlocutor but lack a referring expression for a particular item (Clark & Wilkes-Gibbs, 1986). Of course, the critical question is how humans would behave if they did *not* have recourse to a shared linguistic system, but were in the process of co-creating one. Existing studies in experimental semiotics leave this question unanswered. More promising answers come from observational fieldwork on emerging sign languages, to which we turn next.

1.2 Evidence from observational fieldwork on emerging languages.

The *de novo* emergence of new sign languages affords unprecedented opportunities to observe the actual processes of emergence and conventionalization of referring expressions. The circumstances in Nicaragua are particularly valuable for investigating possible impacts of communicative network structure on the conventionalization of referring expressions.

Two types of young semiotic systems are attested in Nicaragua, each created over comparable periods of time, but arising in social/interactional networks with radically different communicative structure. Some deaf individuals in Nicaragua have acquired Nicaraguan Sign Language (NSL), which is used by ~1,500 people as their primary language. The communicative networks of NSL signers are therefore richly-connected, as schematically illustrated in Figure 1. Meanwhile, other deaf individuals have grown up as "homesigners": lacking access to NSL or any other language, signed or spoken, these individuals have created semiotic systems (homesign) that they use with their immediate interlocutors (friends, family). The communicative network of a homesign system approximates the sparsely-connected network illustrated in Figure 1, where the central node ("A") represents the homesigner, who is connected to each other member of the network but where nodes B, C, and D (e.g. mother, father, sibling, friend) are not connected to each other. This represents the fact that in natural interactions, these people do not use the homesign with each other; they use Spanish instead.

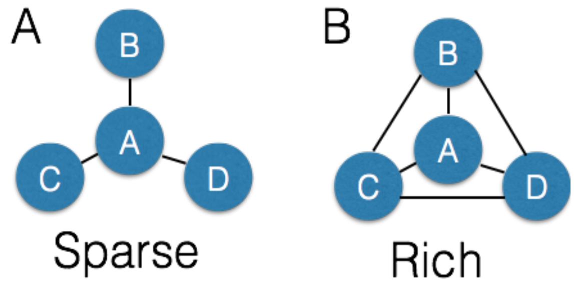

Figure 1

Recently, Richie, Yang, and Coppola (2014) studied the conventionalization of referring expressions among deaf homesigners in Nicaragua and their family/friends, and among NSL signers. They asked each homesigner to produce their referring expression for common items (e.g. cow, rain, sun, egg, etc.), and then did the same with each homesigner's most frequent interlocutors (e.g. parents, siblings, friends). These referring expressions were typically composed of several conceptual components: for example, one person's response to the item 'cow' might be HORNS-MILKING, whereas another's might be MILKING-DRINK. For each item, they then quantified conventionalization by calculating a formal measure of (dis)similarity among the referring expressions produced by members of each homesign family. Critically, conventionalization was weaker among people in homesign networks than among members of the first cohort of Nicaraguan Sign Language users, despite the fact that both homesigners and NSL signers had been using their respective systems for roughly the same amount of time. This is consistent with the hypothesis that richly-connected networks will show greater conventionalization than sparsely-connected networks.

Of course, as with most studies based on naturalistic observation, these findings do not demonstrate a causal impact of communicative network structure on lexical conventionalization. For one thing, even if the difference between homesign and NSL network structures has an impact on conventionalization, the mechanism(s) by which it does so is unclear. For example, in the NSL community, multiple, simultaneous interactions can occur among multiple pairs or groups of interlocutors. In homesign communities, however, only one conversation – with the homesigner involved – can happen at once, because if other members of the homesign community wish to interact they simply speak Spanish. Thus, *per unit time*, the NSL network provides more opportunities for interaction. It may be this, *rather than an advantage in conventionalization per interaction,* that accounts for the greater conventionalization of NSL compared to homesign. Thus, to provide a more rigorous investigation of the causal role of network structure on language emergence, Richie et al. (2014) and others have used computational simulations.

1.3 Evidence from computational simulations.

Because naturally-occurring cases of language emergence are rare and impossible to control, and because of the logistical complexity of conducting experiments with large groups of people interacting over time, computational modeling has emerged as a key tool in studying language emergence generally (e.g., de Boer, 2000; Kirby, Dowman, and Griffiths, 2007) and conventionalization specifically (e.g., Barr, 2004; Puglisi, Baronchelli, & Loreto, 2008). Many of these models have been so-called *agent-based* models (ABMs), as they model how local interaction among multiple agents gives rise to global coordination or structure. Further, agents in ABM's can be situated in arbitrarily complex social networks, making ABM's a powerful tool for understanding the effects of social network structure on the emergence of large-scale structure. Indeed, several studies have used ABM's to understand how social network structure impacts conventionalization of referring expressions (for brief review, see, e.g., the section on 'Simple dynamics on networks' in Baronchelli, Ferrer-i-Cancho, Pastor-Satorras, Chater, & Christiansen, 2013). As mentioned above, Richie et al (2014) supplemented their studies of Nicaraguan Sign Language and homesign with an agent-based model of gesture[2] production, comprehension, and learning. In their model, agents associated gestures (e.g., MILKING, DRINK) to objects ('cow') with certain probability, and sampled those probabilities when attempting to communicate about an object. Listeners likewise attempted to comprehend utterances by producing a string of gestures for the object under discussion (under the assumption that the object could already be disambiguated for both interlocutors by points and other deictics), and comparing their string to the speaker's. Communication succeeded probabilistically as a function of similarity of the speaker's and listener's strings, and, if communication succeeded, the listener rewarded and punished their internal probabilities based on the gestures present or absent in the speaker's utterance. Groups of agents communicating in this fashion eventually conventionalized gestures for particular objects. Critically, however, richly-connected networks of agents conventionalized faster, i*.e. with fewer interactions*, than did sparsely-connected networks, mirroring the differences between homesign and NSL networks in the observational data, and providing a possible mechanistic basis for the network effect.

That result, however, contrasts with that of Gong, Baronchelli, Puglisi, and Loreto (2012). In their model, called the Naming Game, agents carved a perceptual continuum (color) into categories, and then agreed upon labels that refer to one or more categories. In contrast to Richie et al. (2014), where the agents both know the intended referent (e.g., 'cow') but adjust the probabilities of producing corresponding gestures, Gong et al.'s agents must guess the intended referent through their word-referent mappings. Whereas Richie et al. found that rich networks converged much faster than sparse networks, Gong et al. (2012) found that sparse networks offered mostly superior convergence properties compared to rich networks.

Are the Richie et al. (2014) and Gong et al. (2012) results contradictory and in need of resolution? One possibility is that they are simply modeling two different kinds of naturally-occurring conventionalization: one involving multi-morphemic expressions for (again, possibly) pre-existing categories (as in homesign and NSL), and the other involving mono-morphemic expressions developed for co-evolving categories (as in Gong et al., 2012). Another possibility, though, is that one or both models rely on unrealistic assumptions. After all, models can only show what follows from their assumptions, and if the assumptions are wrong, then the model predictions may not be accurate. For example, the agents in Richie et al. (2014) updated their lexicons via reinforcement learning; this is a mode of learning that has wide support in various domains (Niv, 2009), but there is no particular *empirical* evidence that humans use it for lexical conventionalization. Both possibilities point to the need for experiments that test the effects of social network structure on conventionalization with real people, in contexts that are more ecologically valid, while still being amenable to empirical control.

1.4 The present study.

The present study is thus motivated by the fact that despite much previous research, no studies have experimentally tested whether communicative network structure impacts the

---

2 The model is *about* gesture production, comprehension, and learning, but gestures are *represented* by numbers in the model. We make this footnote simply to avoid 'mistaking the map for the territory'.

conventionalization of referring expressions when real humans with real bodies communicate about real things in the real world. We address this gap by randomly assigning naïve human participants to richly- or sparsely-connected networks (approximating NSL and homesign communities, respectively), and asking participants to create referring expressions using gesture alone, without speech. This silent gesture approach combines experimental control with naturalistic validity and yields results that are remarkably consistent across speakers of typologically diverse languages (e.g. Futrell et al., 2015; Goldin-Meadow et al., 2008; Hall et al., 2014; Meir et al., 2017). Converging results from observational fieldwork, computational simulation, and controlled experiments would provide strong evidence that communicative network structure does in fact influence conventionalization of referring expressions. Given the simulation and observational fieldwork results of Richie et al (2014) discussed above, we expected participants in richly-connected networks to conventionalize referring expressions faster than participants in sparsely-connected networks. Further, based on the simulation results, we expected this effect to hold even when controlling for differences between networks in number of interactions among network members.

2 Methods

The Institutional Review Board at the University of Connecticut approved the following methods, to which the participants also gave informed consent.

2.1 Participants.

We asked hearing undergraduates who had no experience with sign language to engage in a dyadic gestural communication task, in groups of four participants each, called "quads". To complete the task, the four members of each quad needed to complete two sessions, typically one week apart. (This time delay was due to the logistical constraints of scheduling 4 participants and 1-2 experimenters, and does not play a substantive role in the theory.) In order to achieve a final dataset of 16 quads (64 individuals), it was necessary to enroll a total of 24 quads (96 individuals). Because partial data could not address the research question, it was necessary to discard entire quads if the data were incomplete. This occurred due to one or more participants in a quad not showing up for or not completing one of the two sessions (n=6 quads), experimenter or equipment error (n=1 quad), or having had previous experience with sign language or gesture, including having previously participated in similar gesture experiments (n=1 quad). Participants received either course credit or $10/session. Initially participants were paid a $5 bonus for completing both sessions; this bonus was later removed when it became clear that it was neither necessary nor effective.

2.2 Design & Procedure

Each quad was assigned to either a sparsely- or richly-connected condition; each participant was also randomly assigned to a position within that network, which we will refer to as A, B, C, and D. Dyads then proceeded to interact as shown in Table 1. Participants within a dyad took turns producing and comprehending gestured descriptions of 25 real-world objects. Speech, writing, mouthing words, and sound effects were prohibited (although participants were allowed to silently *mouth* sound effects like MOO, as this occurs in Nicaraguan homesign, Richie et al., 2014). Each participant had a booklet displaying a target image to describe, as well as an array of 25 images corresponding to the possible items that their interlocutor might describe. The 25 images were the same for both partners, but ordered differently. A low visual barrier occluded the participants' booklets and arrays from their partner's view.

After one participant described an item, the other participant would silently point to one of the items in the 5 x 5 array. The experimenter recorded the selected item, and then said "ok" to indicate that the next trial could begin (the describer and the comprehender switch roles for a new item), regardless of response accuracy.

Round	Room 1 (sparse & rich)	Room 2 (rich only)	Network age, sparse	Network age, rich
1	A-B	C-D	0	0
	A-C	B-C	1	2
	A-D	B-D	2	4
2	A-B	C-D	3	6
	A-C	B-C	4	8
	A-D	B-D	5	10
3	A-B	C-D	6	12
	A-C	B-C	7	14
	A-D	B-D	8	16
4	A-B	C-D	9	18
	A-C	B-C	10	20
	A-D	B-D	11	22

Table 1

After a dyad had described all 25 images to each other, they switched partners. The first "round" was completed once each participant had communicated with all assigned interlocutors. Participants completed rounds 1 and 2 during the first session, and completed rounds 3 and 4 approximately one week later, for a total of four interactions with the same interlocutors. Note that the interactions involving "A" are identical in both conditions.

2.3 Materials

The stimuli were images of 25 items adapted from Richie and Yang (2013), listed in the Appendix. To encourage participants to create expressions that referred to categories (e.g. "girl") rather than specific exemplars (e.g. a particular girl), each item was represented by 3 exemplars. The original Richie and Yang items were created to be recognizable to Nicaraguan participants; we selected a subset of those items that would also be familiar to American college students, and supplemented as necessary to ensure that all items had at least one semantically-similar competitor.

2.4 Coding.

Participants all gave consent to be videorecorded as part of the study. These recordings were used for offline coding by the experimenters. For each of the 14,400 gestured utterances obtained, an experimenter identified the iconically-expressed conceptual components[3]

[3] We focus on conceptual components rather than phonological or articulatory components largely for ease of analysis. In addition, Morgan (2015), in analysis of Kenyan Sign Language and Al-

(e.g., MILKING or EARS for 'cow') that were produced between trial onset and when the interlocutor (correctly or incorrectly) selected an image from the target array. To gauge inter-rater reliability, a second experimenter independently coded 25% of these utterances. Cohen's kappa was computed for each combination of item and quad (e.g., the first rich network's 'avocado' utterances). Average cohen's kappa was .59, indicating moderate, bordering on substantial, agreement, by Landis and Koch's (1977) guidelines.

2.5 Measuring Convergence Across The Network

Measuring convergence across time requires us to choose several sets of utterances, where each set constitutes a snapshot of the network at a particular point in time. One way we could do this is merely considering the utterances from each round. However, this would yield only four time points. A more fine-grained approach is to consider sliding windows over describer-comprehender pairs. In particular, we consider all windows of three consecutive pairs. In the sparse network, this means the first window is AB, AC, AD all from round 1, the second window AC and AD from round 1 and AB from round 2, the third window AD from round 1 and AB and AC from round 2, and so forth. In the rich network, the first window is analogously all the pairs from round 1, the second window the pairs from rows 2 and 3 of round 1 and row 1 of round 2, the third window is row 3 of round 1 and rows 1 and 2 of round 2, and so forth. Constructing windows this way, we obtain 10 time points over which we can measure convergence.

To define the amount of (dis)agreement on gestural conventions within each community, we first encoded each individual utterance as a set of gestures and converted it to a binary vector. Let c, s, t, i, w be particular instances of community (quad), source and target (corresponding to describer and comprehender), item, and (sliding-)window (or network age in number of completed dyads). Let $u(c, s, t, i, w)$ be an utterance that source s produced to target t to refer to item i at window w in community c. Any utterance can be encoded as a binary vector $(f_1, \ldots, f_k, \ldots, f_N)$ where N is the number of unique conceptual components (MILKING, DRINK, etc.; in this section we refer to these as features for brevity) and f_k takes a binary value 0 or 1 depending on whether the k-th feature (or gesture) is absent or present in the utterance.

Let us consider a set of utterances that member s of community c produced to refer to item i at time window w, $U(c, s, i, w) = \{u(c', s', t', i', w') \mid c=c', s=s', i=i', w=w'\}$. For every set $U(c, s, i, w)$, we define a vector $\boldsymbol{p}(c, s, i, w) = (p_1, \ldots, p_k, \ldots, p_N)$ where p_k is the proportion of utterances that have feature k. We take the average of $\boldsymbol{p}(c, s, i, w)$ across all members s, $\boldsymbol{p}(c, i, w) = \sum_s \boldsymbol{p}(c, s, i, w) / 4$. If p_k is 0 or 1, every member of community c completely agrees on whether to use feature k or not to refer to item i at round w and hence entropy is 0. The value of 0.5 suggests the strongest possible disagreement among community members and hence entropy is maximal, at 1. Conditional entropy $H(P_k \mid c, i, w) = - p_k \log(p_k) - (1 - p_k) \log(1 - p_k)$ (we treat $0 \log 0$ as 0)[4] provides a natural way of capturing the amount of uncertainty or disagreement regarding the choice of feature k.

We assume that random variables P_k's are independent from each other once community, item, and round are chosen. Under this conditional independence assumption, we can define the entropy of utterances as follows: $H(U \mid c, i, w) = H(P_1, \ldots, P_k, \ldots, P_N \mid c, i, w) = \sum_k H(P_k \mid c, i, w)$.

Sayyid Bedouin Sign Language, concludes that conventionalization on conceptual components precedes convergence on articulatory components. As we are experimentally observing the very beginnings of a communication system, we find it reasonable to focus on the former stage.

[4] In this paper, we use base 2 for all logarithmic transformation.

Finally, we define the amount of *disagreement* (henceforth *divergence*) of community c at window w as the average of $H(U \mid c, i, w)$ across all items i: $D(c, w) = \mathrm{Avg}_i(H(U \mid c, i, w))$. The range of theoretically possible values of D is $[0, +\infty)$. In our case, all values of $D(c, w)$ were greater than 0.

3 Results

3.1 Accuracy

Accuracy was relatively high even at the beginning of the experiment, averaging around 80% (chance = 4%), and improved to near ceiling by round 4. There were no differences in accuracy as a function of condition, confirming that both sets of participants understood and complied with task instructions. Figure 2 shows only accuracy for dyadic interactions that were common to both conditions (Table 1, shaded region); accuracy for the interactions unique to the Full condition was similarly high.

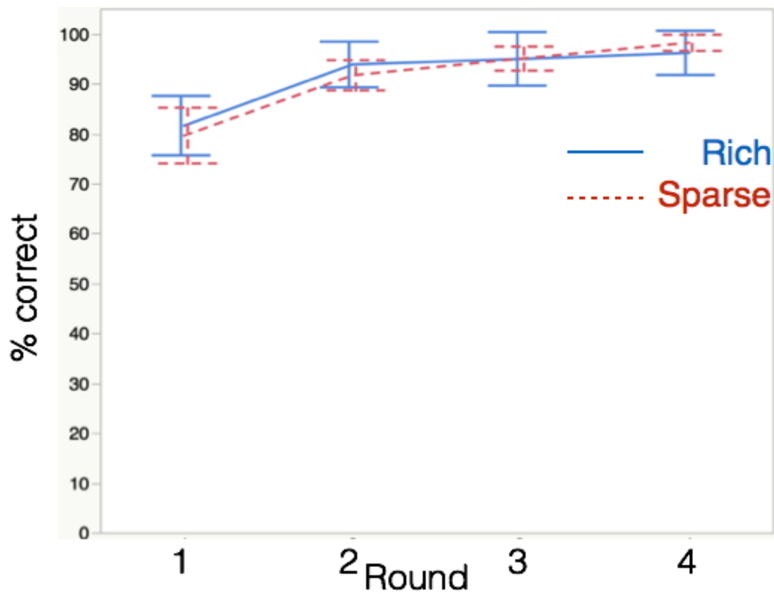

Figure 2

3.2 Conventionalization

The visual inspection of data (see thin lines in Figure 3) suggests that the logarithm of divergence reduced approximately linearly as a function of SlidingWindow (or simply Window). We used `brms` (Bürkner, 2017), a Bayesian multilevel modeling package, to fit log(divergence) to a Bayesian linear mixed-effects model that includes Window and Window-by-NetworkType interaction as fixed-effects terms as well as by-Quad random intercepts, random slopes of Window, and their correlations. We did not include the fixed-effect term of NetworkType because log(divergence) is expected to be the same when Window = 0 when the members of a community had not interacted with each other yet. The categorical variable NetworkType was effect-coded such that the contrasts of the Full and Star network conditions were set to 0.5 and -0.5, respectively. Thus, the coefficient of the interaction term Window-by-NetworkType is interpreted as the difference between the rate of change in the Full networks and the rate of change in the Star networks.

Table 2 presents the estimates of the coefficients of (Sliding)Window and Window-by-NetworkType with 95% credible intervals. The coefficient of Window was negative and did not include 0, b_{Window} = -0.138, 95% CI = [-0.159, -0.117], suggesting that log(divergence) reduced as community members participated in more communicative interactions across time. More importantly, the slope difference between Full and Sparse networks was negative and did not include 0, b_{WxNT} = -0.047, 95% CI = [-0.086, -0.008], suggesting that log(divergence) decreased more quickly with Window in the Full networks than in the Star networks. Figure 3 presents observed divergence values as well as the fitted values with 95% credible intervals on the log divergence scale.[5] Although we prefer the parameter estimation approach to the Bayesian hypothesis testing approach, we report the Bayes factor for those who are interested. We calculated the Bayes factor to see whether the reported model (H1) with the interaction term is favored over a simpler model (H0: $b_{WxNT} = 0$) without the interaction term and observed BF_{10} = 2.387, suggesting that, *under our choice of the prior*, there is no strong evidence favoring H1 over H0, although the direction is as expected[6]. We briefly note that the frequentist analysis of the same data suggests that all the coefficients (including the interaction term) were statistically significant; in particular, b_{AxNT} = -0.048, SE_{WxNT} = 0.019, $t(14.0) = -2.55$, $p = .023$.[7] More detailed information including our choice of prior distributions can be found in the Supplementary Materials.

Table 2. Summary of the model estimates.

	Estimate	Est.Error	95% CI
Fixed-effects			
Intercept	2.170	0.074	[2.022, 2.314]
Window	-0.138	0.011	[-0.159, -0.117]
Window x NetworkType	-0.047	0.020	[-0.086, -0.008]
Random-effects			
SD(Intercept)	0.284	0.063	[0.191, 0.436]
SD(Window)	0.040	0.009	[0.026, 0.062]
Corr(Intercept, Window)	-0.141	0.255	[-0.606, 0.376]

[5]Figure 3 suggests that the divergence of one Quad (S09) in the Star network condition was exceptionally large. Because we observed divergence reduction in the quad suggesting the quad followed the experimental instruction well, we think the group should not be excluded but be explained by the model. For those who are interested, however, we report the result from the analysis after excluding S09 here: b_W = -0.137, SE_W = 0.011, 95% CI = [-0.159, -0.114]; b_{WxNT} = -0.046, SE_{WxNT} = 0.023, 95% CI = [-0.0909, -0.0003].

[6]Our vague prior for the NetworkType-by-Window interaction effect assigns some probability mass to unlikely large values, which are expected when the divergence reduces at the fastest rate in one condition but increases at the fastest rate in another condition. We suspect this vague prior penalizes H1 too strongly.

[7]The same model was fitted to data, using the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015). Following Luke (2017), we computed *p*-values based on the Kenward-Roger approximation of the degrees of freedom implemented in the `lmerTest` package (Kuznetsova, Brockhoff, & Christensen, 2017).

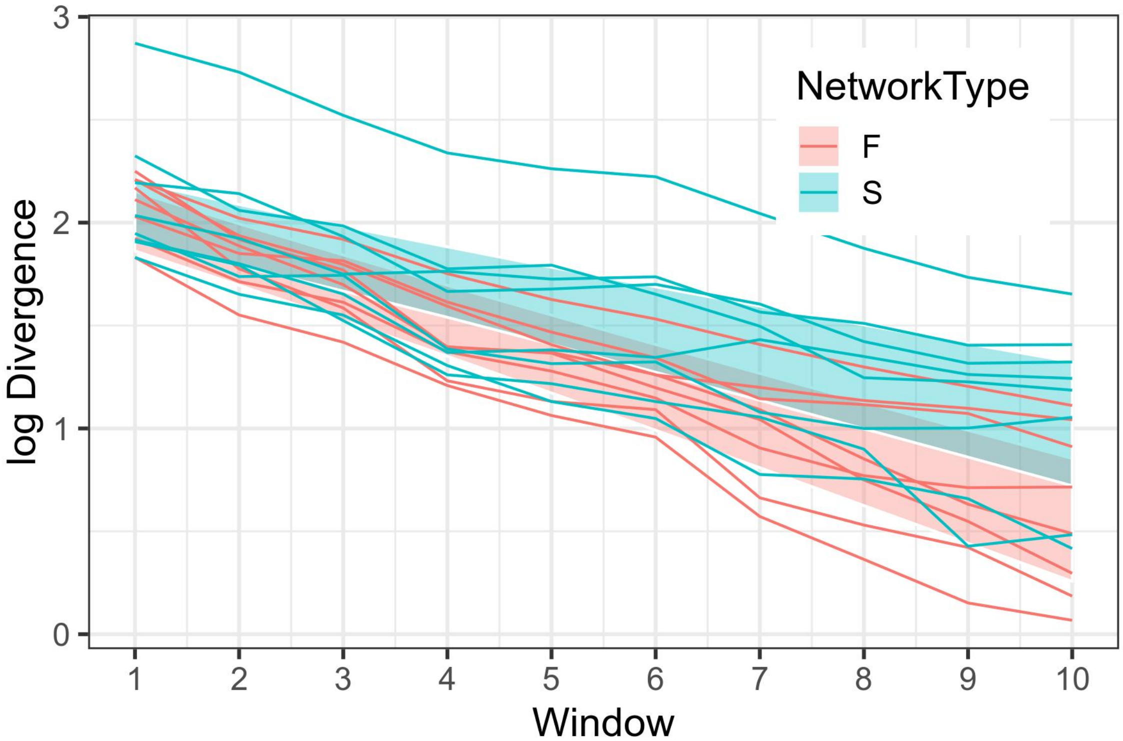

Figure 3.

3.2.1 Network Age

A byproduct of our manipulation of network structure is that twice as many communicative interactions occur per round in the rich condition, relative to the sparse condition. We refer to this as "network age". Therefore, one possibility is that there is no effect of network *structure* beyond that of network *age*. To evaluate this alternative interpretation, we must measure the degree of conventionalization at time points where the two conditions are equated for network age. If network age explains the difference between the conditions, there should be no difference between the groups after equating for network age; if a difference in the same direction persists, it will constitute evidence of an impact of network structure that cannot be explained by network age.

To control the number of interactions at the community level, we refit the data to the same model; the only difference was the replacement of predictor Window with NetworkAge (NA).[8] Table 3 presents the estimates of the model parameters as well as their credible intervals. Interestingly, and contrary to predictions, when the number of interactions within each community was controlled, the communities of the Full network type converged on gestural conventions more slowly than the communities of the Star network type; $b_{NAxNT} = 0.032$, $SE_{NAxNT} =$

8 We note that for two network types, we observed different ranges of values in NetworkAge: 1-10 in the Star network type and 2-20 in the Full network type. This is not problematic when we use a linear model as we do in the present study, but fitting this dataset to nonlinear models would present certain risks.

0.015, 95% CI = [0.002, .061].[9] The Bayes factor (under our choice of priors) favors a simpler model without the interaction, although it is not meaningful: $BF_{10} = 1.134$. The frequentist analysis suggests that the coefficient of the interaction was statistically significant; $b_{NAxNT} = 0.033$, $SE_W = 0.014$, $t(14.1) = 2.3$, $p = .037$. More detailed information is presented in the Supplementary Materials.

Table 3. Summary of the model estimates.

	Estimate	Est.Error	95% CI
Fixed-effects			
Intercept	2.171	0.074	[2.024, 2.316]
NetworkAge	-0.098	0.008	[-0.114, -0.082]
NetworkAge x NetworkType	0.032	0.015	[0.002, 0.061]
Random-effects			
SD(Intercept)	0.284	0.062	[0.190, 0.431]
SD(NetworkAge)	0.030	0.007	[0.020, .047]
Corr(Intercept, NetworkAge)	-0.158	0.256	[-0.615, 0.365]

9 When the quad (S09) with high divergence was excluded and then the data was fitted to the model, we observed a different pattern: $b_{NA} = -0.067$, 95% CI = [-0.079, -0.056]; $b_{NAxNT} = 0.023$, 95% CI = [-0.001, 0.046], suggesting that the convergence rate (the absolute value of the slope) was not different between two network types.

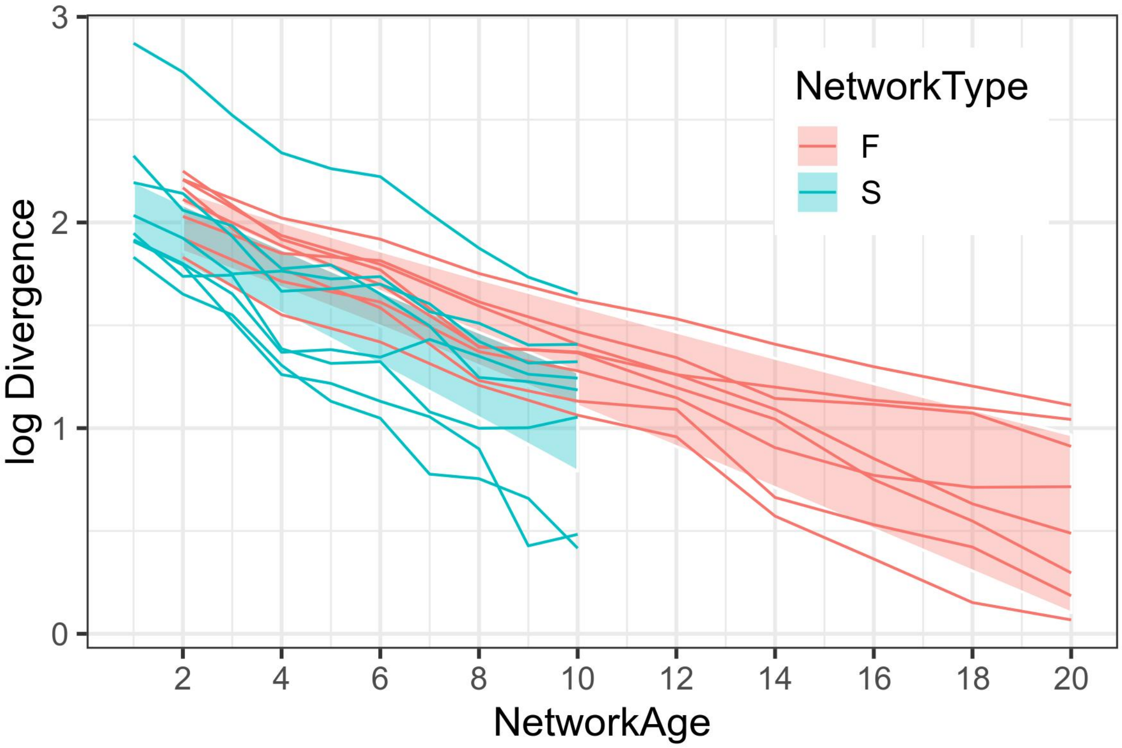

Figure 4

4 Discussion

We began by asking how people agree on what to call things. By randomly assigning naïve participants to conditions that differed only in communicative network structure, we discovered that, *per unit time*, conventionalization is greater and grows more quickly in the rich condition, where all possible pairs communicate with one another and *multiple pairs can communicate independently and simultaneously*, relative to the sparse condition, where all communication is channeled through a central hub, limiting the network to one interacting pair at a time. However, we found that this condition difference *reversed* when controlling for rich networks' affordance of parallel interactions, suggesting that, *per interaction*, the sparse network conventionalizes faster. The first finding is consistent with observational fieldwork from a host of gestured and signed systems (Richie et al., 2014; Horton, Goldin-Meadow, & Brentari, 2016), but the latter finding contradicts predictions made from a previous agent-based model of conventionalization (Richie et al., 2014).

This difference between the networks that are optimal for rate of conventionalization *per unit time,* and rate of conventionalization *per interaction,* suggests a subtler account of network effects on language emergence than initially supposed (Richie et al., 2014). That more interactions hastens conventionalization is likely fairly obvious – more interactions allow for more influence of communication partners on one another. But what accounts for the surprising finding that sparse networks conventionalize faster *per interaction* than rich networks? A few possibilities present themselves. First, perhaps in sparse networks, the hub (participant A) can set a standard, and then the rest of the network 'merely' needs to converge on them. In rich networks, however, members must in some sense negotiate whose utterances they will converge on, presenting a symmetry-breaking problem of some form. Along similar lines, if every participant is trying to remember something about every other participant's preferred utterances, overall memory demands will be higher in the full network, potentially leading to degraded memory and convergence on other members' preferred referring expressions as a way of collectively reducing the amount of information that needs to be encoded, stored, and retrieved. Ideally, any account of present network structure effects would be implemented and tested in an agent-based model. Richie (2017) presents one such newer model of the conventionalization process in the present setting, but predictions about network effects have yet to be generated from this model, leaving this one avenue for future research.

However, caution in overgeneralizing the present results is warranted. First, comprehenders received no feedback on whether they chose (in)correctly. High communication accuracy and conventionalization were nevertheless possible because the target items afforded highly iconic and unambiguous utterances. In more naturalistic settings, of course, interlocutors often have the opportunity to detect and rectify misunderstandings. It is unclear how this difference might affect the present results and conclusions, although one possibility is that allowing feedback would merely hasten conventionalization and improvement in communication and not otherwise influence network effects. Second, the present networks are simplifications relative to real world linguistic communities. For example, real Nicaraguan home sign networks are larger than 4 individuals, and in the real Nicaraguan Sign Language community, it is almost certainly not the case that every member interacts with every other member with equal frequency. The NSL community may well be characterized more accurately by something like a small-world network (Watts and Strogatz, 1998), a network with mostly local connections and a few random long-range connections. An experimental study of conventionalization would of course have more ecological validity if the experimental networks more closely matched those of the analogous communities in the real world, but running an experiment in the gestural modality at greater scale and complexity comes with significant practical challenges. It thus may be more feasible to run web-based experiments in a nongestural modality but with more realistic network structures (e.g., Centola & Baronchelli, 2015)[10]. Of course, this would entail sacrificing the affordances of the human body, physical environment, and knowledge about the affordances of the referents that silent gesture studies offer. Ideally, combining studies using naturalistic communication media like the present and results from studies using more realistic network structures will provide convergent results.

Experimentally testing and simulating conventionalization in more complex settings is also warranted as the relationship between increased connectedness – and consequently opportunity for parallel interaction – and increased conventionalization may not be monotonic. For example, factors that increase conventionalization within a specific network could subsequently impede conventionalization if and when members of different networks attempt to communicate with one another. People from richly-connected networks may have *overlearned* the referring expressions internal to their own network, such that they may not conventionalize with new interlocutors as well as people from more sparsely-connected networks might.

---

10 Besides experimentally testing networks motivated by particular naturally occurring languages and communication systems, larger web-based experiments can also manipulate particular network structural properties, like node degree, clustering, average path length, modularity, etc (Baronchelli et al., 2013).

Observational data from Al-Sayyid Bedouin Sign Language (ABSL; Sandler et al., 2011) and Kenyan Sign Language (Morgan, 2015) are consistent with this possibility. As described by Sandler et al. (2011), ABSL is a full human language that lacks a system of phonological contrasts. More precisely, linguists studying ABSL have thus far not been able to identify minimal pairs: lexical items whose forms differ from one another only with respect to meaningless formational elements (e.g. handshape, movement, location). In addition, they find that the phonological form of many common referring expressions differs slightly from family to family. They propose that patterns of social interaction partially explain this pattern: family members likely interact more with one another than with other users of ABSL, and conventionalization at this relatively local level may impede conventionalization across the entire community, and/or may function as a sociolinguistic identity marker. Kenyan Sign Language, on the other hand, despite being about 20 years younger than ABSL, provides clear and compelling evidence for a system of phonological contrasts (Morgan, 2015)[11]. Might differences in the social structure of these communities account for these differences in lexical conventionalization? KSL has a much larger group of signers, and is a classic example of a "Deaf community sign language" whereas ABSL is a classic example of a "village sign language", as per Meir et al. (2010). Unfortunately, too many uncontrolled factors differ between ABSL and KSL to offer a definitive answer to this question. However, other research on emerging sign languages affords even more promising answers.

Horton, Goldin-Meadow, and Brentari. (2016) have begun to document the process of lexical conventionalization among deaf homesigners living in Guatemala. Whereas most research on homesigners focuses on those who are the only deaf member of their family, Horton's work is remarkable for its inclusion of homesigners who have *other deaf homesigners* as interlocutors: whether siblings, extended family members, or even parents (see also Haviland, 2013, 2015). Because these individuals are embedded within relatively similar cultures (e.g. they all live near Nebaj, Guatemala, cf. Nicaraguan homesigners), differences in conventionalization among these individuals are more likely to reflect the impact of the different communicative network structures. She finds that when there are multiple homesigners in a family, conventionalization within that family is intermediate between what is observed in the families of individual homesigners and in sign language communities.

In each of these naturally-emerging languages, conventionalization for referring expressions is greater in richly-connected networks than sparsely-connected networks. Although we could not infer causality on the basis of these observational data alone, the present study obtained an analogous pattern via random assignment in an experimental context. Again, our results suggest this rich network advantage arises not because rich networks provide greater conventionalization *per interaction* (in fact, quite the opposite holds), but because more richly-connected networks afford more communicative interactions per unit time (as multiple conversations can happen at once), which is virtually certain to be the case in real-world rich

---

[11] Morgan (2015) further observes that conventionalization seems to proceed in at least two stages. Whereas the examples described from ABSL mostly fall within what Sandler et al. (2011) identify as an "iconic prototype", Morgan (2015) points out that there is likely to have been a separate, earlier phase during which signers first converge on a "conceptual target". She finds that conventionalization toward an "articulatory target" tends to follow at a slight lag, suggesting that conventionalization begins at a semantic level before continuing at a phonological level. Conventionalization at this semantic level is therefore the primary focus of the present study.

networks (e.g., networks with multiple Deaf individuals carrying on multiple independent signed conversations).

Here, we have restricted ourselves to referring expressions, which we take to be an elemental ingredient of any emerging language, including homesign. Even under a generative view, structure at this level is expected to depend on external input; therefore, the present findings do not pose any particular challenges to theories of linguistic structure as conceived under the minimalist program (Chomsky 1995). Still, they do highlight the potential for functional or usage-based influences on the lexicon (Piantadosi, Tily, and Gibson, 2011; Richie, 2016), as well as lead to consideration of the potential roles that the lexicon may or may not play in the emergence of other aspects of linguistic structure. For example, the emergence of structure at discourse and perhaps even syntactic levels seems to not require a conventionalized lexicon (e.g., Coppola & Newport 2005); indeed, even hearing non-signers are highly consistent in their use of constituent order when describing various types of events (Christensen et al., 2016; Futrell et al., 2015; Gibson et al., 2013; Goldin-Meadow et al., 2008; Hall et al., 2013, 2014, 2015; Meir et al., 2017; Schouwstra & de Swart, 2014). Interestingly, hearing non-signers show much less systematicity at morphophonological levels (Brentari et al., 2012, 2015, 2016; Horton et al., 2015).

Finally, we acknowledge that those who favor hypothesis testing with Bayes factors over parameter estimation with credible intervals – which we favor and interpret here – may reach different conclusions from our data. Specifically, Bayes factors suggested there is no strong evidence for Network-Type by Window or Network-Type by Network-Age interactions. We are hesitant to make a conclusion based on these Bayes factors because, for one, the null model instantiates an interaction effect of precisely 0, which is highly unlikely for almost any effect (Gelman, 2011) and especially in our case where participants in different Network Type conditions were exposed to different communication environments. We also note that our priors were vague, assigning a significant proportion of probability mass to unlikely large interaction effects, due to lack of similar previous work; as well known, the Bayes factor can be highly sensitive to the choice of priors. Regardless of one's preferred interpretation of the present findings, the data reported here can better inform the priors for future work.

5 Conclusions

The study of homesign and emerging sign languages sheds new light on the earliest stages of semiotic organization in human languages. Compared to previous results from observational fieldwork and computational simulation, our experimental data suggest a more subtle story concerning the role of network structure in language emergence. We have demonstrated that people in more richly-connected networks tend to converge on referring expressions more quickly and more completely than those in sparsely-connected networks, although, contrary to previous computational modeling, this difference reverses when controlling for the greater numbers of interactions afforded by rich networks. This latter result points to the importance of testing computational models of language emergence against empirical data. Besides using present posteriors to justify informative new priors in replicating the present work in a Bayesian data analysis framework, future experimental and modeling work, inspired by observational fieldwork, could investigate effects of pseudo-lexical conventionalization on emergence of other linguistic structures, and conventionalization (and its consequences) in richer, more realistic networks.

Acknowledgements

This research was supported by NSF IGERT Grant 1144399 to the University of Connecticut, and the second author was supported by NIH 5F32DC013239. We are particularly indebted to Sarah Lodge, Megan Brown, Jenna Shaw, and Shea Marsella for their tireless efforts in collecting and coding these data, as well as to colleagues at EvoLang 11 and several anonymous reviewers of previous versions of this manuscript.

References

Baronchelli, A., Felici, M., Loreto, V., Caglioti, E., & Steels, L. (2006a). Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics: Theory and Experiment*, *2006*(06), P06014.

Baronchelli, A., Dall'Asta, L., Barrat, A., & Loreto, V. (2006b). Topology-induced coarsening in language games. *Physical Review E*, *73*(1), 015102.

Baronchelli, A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in Cognitive Sciences*, *17*(7), 348-360.

Barr, D. J. (2004). Establishing conventional communication systems: Is common knowledge necessary?. *Cognitive Science*, *28*(6), 937-962.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482.

Brentari, D., M. Coppola, P. W. Cho, and A. Senghas. (2016). Handshape complexity as a precursor to phonology: Variation, emergence, and acquisition. *Language Acquisition.* DOI:10.1080/10489223.2016.1187614

Brentari, D., Coppola, M., Mazzoni, L., & Goldin-Meadow, S. (2012). When does a system become phonological? Handshape production in gesturers, signers, and homesigners. *Natural Language & Linguistic Theory*, *30*(1), 1-31.

Brentari, D., Renzo, A. D., Keane, J., & Volterra, V. (2015). Cognitive, cultural, and linguistic sources of a handshape distinction expressing agentivity. *Topics in Cognitive Science*, *7*(1), 95-123.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1-28. doi:10.18637/jss.v080.i01

Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences*, *112*(7), 1989-1994.

Chomsky, N. (1995). *The Minimalist Program* (Vol. 28). Cambridge, MA: MIT Press.

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

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

Coppola, M., & Newport, E. L. (2005). Grammatical subjects in home sign: Abstract linguistic structure in adult primary gesture systems without linguistic input. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(52), 19249-19253.

Coppola, M., & Senghas, A. (2010). The emergence of deixis in Nicaraguan signing. In D. Brentari, (Ed.), *Sign languages: A Cambridge Language Survey*, 543-569.

Dall'Asta, L. et al. (2006) Agreement dynamics on small-world networks. Europhys. Lett. 73, 969.

de Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, *28*(4), 441–465.

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

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, *34*(3), 351-386.

Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., & Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition*, *136*, 215-221.

Galantucci, B., Garrod, S., & Roberts, G. (2012). Experimental semiotics. *Language and Linguistics Compass*, *6*(8), 477-493.

Gelman, A. (2011). So-called Bayesian hypothesis testing is just as bad as regular hypothesis testing. Retrieved from https://statmodeling.stat.columbia.edu/2011/04/02/so-called_bayes/

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word order variation. *Psychological Science, 24*(7), 1079-1088.

Goldin-Meadow, S., Butcher, C., Mylander, C. & Dodge, M. (1994). Nouns and Verbs in a Self-Styled Gesture System: What's in a Name? *Cognitive Psychology, 27*, 259-319.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*(27), 9163-9168.

Gong, T., Baronchelli, A., Puglisi, A., & Loreto, V. (2012). Exploring the roles of complex networks in linguistic categorization. *Artificial Life*, *18*(1), 107-121.

Hall, M. L., Ahn, Y. D., Mayberry, R. I., & Ferreira, V. S. (2015). Production and comprehension show divergent constituent order preferences: Evidence from elicited pantomime. *Journal of Memory and Language*, *81*, 16-33.

Hall, M. L., Ferreira, V. S., & Mayberry, R. I. (2014). Investigating constituent order change with elicited pantomime: A functional account of SVO emergence. *Cognitive Science*, *38*(5), 943-972.

Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*(1), 1-17.

Haviland, J. B. (2013). The emerging grammar of nouns in a first generation sign language: specification, iconicity, and syntax. *Gesture*, *13*(3), 309-353.

Haviland, J. B. (2015). Hey!. *Topics in Cognitive Science*, *7*(1), 124-149.

Horton, L., Goldin-Meadow, S., & Brentari, D. (2016). Homesign Contact and Conventionalization of a Lexicon. Talk presented at the 41st BUCLD Conference, Boston MA.

Horton, L., Goldin-Meadow, S., Coppola, M., Senghas, A., & Brentari, D. (2015). Forging a morphological system out of two dimensions: Agentivity and number. *Open Linguistics*, *1*(1).

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30-66.

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, *104*(12), 5241-5245.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Meir, I., Aronoff, M., Börstell, C., Hwang, S. O., Ilkbasaran, D., Kastner, I., ... & Sandler, W. (2017). The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition*, *158*, 189-207.

Morgan, H. E. (2015). When does a word emerge? Lexical and phonological variation in a young sign language. San Diego Linguistic Papers, 5, 2-17. Retrieved from: https://escholarship.org/uc/item/0dg64753

Niv, Y. (2009). Reinforcement learning in the brain. *The Journal of Mathematical Psychology, 53*(3), 139–154.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences, 108*(9), 3526-3529.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02), 169-190.

Puglisi, A., Baronchelli, A., & Loreto, V. (2008). Cultural route to the emergence of linguistic categories. *Proceedings of the National Academy of Sciences*, *105*(23), 7936-7940.

Richie, R. (2016). Functionalism in the lexicon: Where is it, and how did it get there? In G. Jarema, G. Libben, & V. Kuperman (Eds.), Thematic Issue of *The Mental Lexicon: New Questions for the Next Decade*, *11*(3), 429–466. Amsterdam: John Benjamins. doi 10.1075/ml.11.3.05ric.

Richie, R., & Yang, C. (2013, July). Elicitation and analytic methods for documenting the lexicons of emerging signed languages. Poster presented at the 11th Theoretical Issues in Sign Language Research, London, England.

Richie, R., Yang, C., & Coppola, M. (2014). Modeling the emergence of lexicons in homesign systems. *Topics in Cognitive Science*, *6*(1), 183-195.

Richie, R., M. Coppola, and C. Yang. (2014). Emergence of natural language lexicons: Empirical and modeling evidence from Homesign and Nicaraguan Sign Language. In W. Orman and M. J. Valleau*, (Eds.), Proceedings of the 38th Annual Boston University Conference on Language Development, pp.* 355-367.

Sandler, W., Aronoff, M., Meir, I., & Padden, C. (2011). The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory, 29*(2), 503–543.

Saussure, F. d. (1983) *Course in General Linguistics*. Eds. C. Bally and A. Sechehaye. (R. Harris, Trans.). La Salle, Illinois: Open Court. (Original work published 1916).

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431-436.

Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science*, *12*(4), 323-328.

Spike, M., Stadler, K., Kirby, S. and Smith, K. (2016). Minimal requirements for the emergence of learned signaling. *Cognitive Science*. doi:10.1111/cogs.12351

Tamariz, M. (2017). Experimental studies on the cultural evolution of language. *Annual Review of Linguistics*, *3*, 389-407.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’networks. *Nature, 393*(6684), 440.

Appendix

Stimulus Items

avocado
baseball cap
beans
boy
cabbage
cloud
cow
cowboy hat
dog
girl
goat
horse
lake
lime
man
mountain
old woman
orange
pig
policeman
rice
sheep
soldier
truck
woman

Tables

Table 1. Schedule of interactions. The sequence of dyadic interactions is given below. Note that in the Rich condition, dyads in Room 2 were tested at the same time as dyads in Room 1; therefore, the Rich network "ages" twice as quickly as the Sparse network.

Table 2. Summary of the mixed-effects regression model. logDivergence ~ 1 + NetworkType + NetworkType:Window + (1 + Window | Quad)

Table 3. Summary of the mixed-effects regression model. logDivergence ~ 1 + NetworkType + NetworkType:NetworkAge + (1 + NetworkAge | Quad)

Figures

Figure 1. Two networks of four individuals each, differing only in connectivity.

Figure 2. Average accuracy across dyads by round, for interactions common to both Rich and Sparse networks.

Figure 3. Plot of log divergence against sliding window. Thin lines represent observed log divergence change. The ribbons present the 95% credible intervals of the conditional means estimated from the model.

Figure 4. Plot of log divergence against network age. Thin lines represent observed log divergence change. The ribbons present the 95% credible intervals of the conditional means estimated from the model.