THE IMPACT OF COMMUNICATIVE NETWORK STRUCTURE ON THE CONVENTIONALIZATION OF REFERRING EXPRESSIONS IN GESTURE

MATTHEW HALL[1,2], RUSSELL RICHIE[2,3,4], MARIE COPPOLA[1,2,3]

[1] *Department of Linguistics*

[2] *IGERT in Language Plasticity*

[3] *Department of Psychological Sciences*

[4] *The Connecticut Institute for the Brain and Cognitive Sciences*

University of Connecticut,
Storrs, CT, USA

matthew.l.2.hall@uconn.edu, russell.richie@uconn.edu, marie.coppola@uconn.edu

The emergence of referring expressions is a critical component of the evolution of any linguistic system. Building on evidence from naturally emerging sign languages as well as computational simulations, we use a behavioral experiment to investigate how the structure of a communicative network influences the processes of conventionalization of referring expressions. We asked hearing individuals who do not have experience with a sign language to engage in a gestural communication task, and randomly assigned them to either a sparsely connected or richly connected network. Pairwise conventionalization was consistent in both conditions, but network-wide conventionalization was greater in the richly connected network. This is the first time this effect has been demonstrated in a controlled experiment in which humans communicate in a natural linguistic modality (i.e. gesture). Differences in the number of communicative interactions may account for the network effect in the present data; results in the literature are mixed on this point.

1. Introduction

1.1. *Conventionalization*

The role of social convention in creating referring expressions (i.e. names for things) has been a question of interest since as early as Plato. Within modern scholarship, there have been observational studies of emerging languages, computational simulations, and controlled experiments in various non-linguistic modalities (for reviews, see Meir et al., 2010; Steels, 2011, and Galantucci et al., 2012, respectively). However, there is a curious lack of experimentation on how referring expressions emerge in communities and modalities in which new languages are known to arise: that is, small numbers of human beings communicating in the manual modality (silent gesture). By linking perceptual form to conceptual meaning, referring expressions are arguably the initial entry point into language structure. Only once they are in place (even if not fully conventionalized) can we begin to observe the emergence of other kinds of structure (e.g. syntax, phonology).

1.2. *Referring Expressions*

Imagine that you wish to communicate the notion of an avocado, but neither you nor your interlocutor has a word for it, despite having knowledge about it. One natural step would be for you to communicate the concept by referencing your knowledge about it as a string of semantic features: for example, its physical properties and affordances, customary uses, emotional valence, etc. Such a string would be the beginning of a referring expression. Objects in the world have a great variety of features that could be listed, of which any given speaker will choose only a finite number (e.g. "it's kinda oblong, it's green, and you eat it"). Moreover, it is likely that your interlocutor might initially choose different features (e.g. "you slice it, take out the pit, and eat it"). This type of communication can be effective to a degree, but is a far cry from the lexicalized symbol "avocado". By what processes do these initially idiosyncratic referring expressions come to be shared, not only by a pair of interlocutors, but by a broader community?

As different speakers interact, they might each retain their own preferred expression, which would minimize the cognitive burden on the producer, but risks being communicatively ineffective, or at least inefficient, if the perceiver's preferred expression is different. Alternatively, if a hypothetical producer wants to communicate with a familiar perceiver, she could retrieve that individual's preferred expression from memory. This might increase communicative success but would (presumably) be representationally costly, especially as the number of potential interlocutors and potential referents grows. Another possibility is for our producer to allow her own mental proto-lexicon to be updated by the various interactions she has with multiple interlocutors. She may observe an innovation that is particularly effective, or that some features are becoming more common than others, and then update her own preferred referring expression accordingly. A population of such agents will eventually converge on a stable mapping of forms, in terms of which semantic features are represented (Richie et al., 2014). Along the way, reduction may also happen, in which excess verbiage is pruned away and the iconic links between form and meaning give way to perceptual/motor ease, increasing apparent arbitrariness in the signal and perhaps enabling the emergence of phonology in the traditional sense.

1.3. *Extant Data*

Processes resembling the above have been reported in at least two emerging sign languages: Al-Sayyid Bedouin Sign Language (ABSL; Sandler et al., 2005) and Kenyan Sign Language (KSL; Morgan, 2015). Referring expressions exist (and have existed for decades) in both languages, but they have not completely conventionalized throughout the community, at the level of either the semantic features represented or in their phonological form. Interestingly, Nicaraguan Sign Language (NSL), which is even younger than ABSL and KSL, appears to have fully conventionalized its repertoire of basic referring expressions (Richie et al., 2014). What might account for the varying speed at which these naturally-emerging systems conventionalize?

1.4. *Communicative Network Structure*

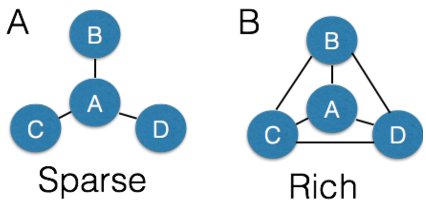

Figure 1. Schematic illustration of sparsely- and richly-connected networks.

Richie et al. (2014) proposed that one important factor might be the structure of the communicative network. They compared users of homesign, in which a single deaf individual (Fig. 1A, center) uses gestural referring expressions to communicate with their hearing friends and family (who use spoken language, rather than homesign, when communicating with one another; Fig. 1A, periphery) to users of NSL, who all use NSL to communicate with each other (Fig. 1B). Using computational simulations, they showed that richly connected networks (NSL-like) do in fact conventionalize faster than sparsely connected networks (homesign-like). However, such naturalistic and computational results are largely suggestive. The naturalistic data contain too many confounding variables to single out network structure as a causal factor, and the computational model only shows what obtains given certain (well-motivated) assumptions put into the model. Thus, it remains to be seen whether actual human beings, in a carefully controlled experimental setting, will show a similar pattern when randomly assigned to a richly- or sparsely connected network. This experimental approach also allows us to control for other relevant sociocommunicative variables that differ between homesign and NSL.

Table 1. Sequence of dyadic interactions in the two conditions.

Round	*Sparse Network*	*Rich Network*	
1	**A-B**	**A-B**	C-D
	A-C	**A-C**	B-D
	A-D	**A-D**	B-C
2	**A-B**	**A-B**	C-D
	A-C	**A-C**	B-D
	A-D	**A-D**	B-C
3	**A-B**	**A-B**	C-D
	A-C	**A-C**	B-D
	A-D	**A-D**	B-C
4	**A-B**	**A-B**	C-D
	A-C	**A-C**	B-D
	A-D	**A-D**	B-C

1.4 *Present Study*

We asked hearing undergraduates who had no experience with sign language to engage in a dyadic gestural communication task. Fourteen groups of four naïve participants each were randomly assigned to either a sparsely- or richly-connected condition; each participant was also assigned a "letter" within that condition, indicating their position in the network: A, B, C, or D. Dyads then proceeded as shown in Table 1. Participants took turns producing and comprehending gestured descriptions of real-world objects. Each participant had a booklet displaying a target stimulus to describe, as well as an array of 25 images corresponding to the possible items that their interlocutor might describe. The 25 images were identical for both partners, but ordered differently.

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 two rounds on day 1, followed by two additional rounds approximately one week later, for a total of four interactions with the same interlocutors. Table 1 shows the sequence of dyads in both conditions. Note that the interactions involving "A" (in bold) are identical in both conditions. These form the core of our analysis.

For each description, we coded the semantic features that were produced between trial onset and when the interlocutor (correctly or incorrectly) selected an image from the target array. We measured conventionalization by computing the similarity between strings of semantic features. The similarity between two unordered strings was quantified by the Jaccard index: the ratio of their intersection to their union. For example, the strings {a, b, c, d, a} and {c, d, x} have a Jaccard index of 2/5 (0.4), because their intersection contains 2 unique elements (c, d) and their union contains 5 unique elements (a, b, c, d, x). Thus, a

Jaccard index of 0 reflects complete divergence, while 1 reflects complete convergence, regardless of length.

2. Results

2.1. *Comprehension accuracy.*

Despite having no previous experience with sign language or pantomime, participants communicated successfully most of the time. Accuracy increased over time, as expected, from means of 83% and 78% after Round 1 to means of 96% and 98% after Round 4 (for rich and sparse conditions, respectively). There was no main effect of condition [$F(1,12) = .15$, $p = .70$], but there was a marginally significant interaction between condition and round[$F(1,12) = 4.51$, $p = .06$], with the sparsely connected network showing greater improvement from Round 1 to Round 4.

2.2. *Direct comparisons: Pairwise conventionalization*

Table 2. Direct comparisons measure dyadic conventionalization. Indirect comparisons measure network-wide conventionalization.

Direct	*Indirect*
A-to-B vs. B-to-A	B-to-A vs. C-to-A
A-to-C vs. C-to-A	B-to-A vs. D-to-A
A-to-D vs. D-to-A	C-to-A vs. D-to-A

To measure conventionalization over time within any given dyad, we plot the average Jaccard index for that dyad from Round 1 to 4 (Table 2 & Figure 2, left). Unsurprisingly, conventionalization increases substantially [$F(1,12) = 247.41$, $p < .001$]. We also found no significant difference between the richly- and sparsely-connected networks [$F(1,12) = .33$, $p = .57$]. Note, however, that under this analysis, high values could reflect dyads establishing a convention that was unique to them as a pair, rather than a network-wide convention.

2.3 Indirect comparisons: Network-wide conventionalization

To assess the extent to which conventionalization increased in the network *as a whole*, we measure the similarity of strings that were *not* produced in direct communication, but rather in indirect communication (Table 2 & Figure 2, right). Specifically, we compare the similarity of the strings produced to participant A by participants B, C, and D. Increases in this measure cannot be due to partner-specific effects, but must reflect increased conventionalization throughout the network. Here, we again find a general increase across rounds in both conditions [$F(1,12) = 197.79$, $p < .001$]. However, we now find that the

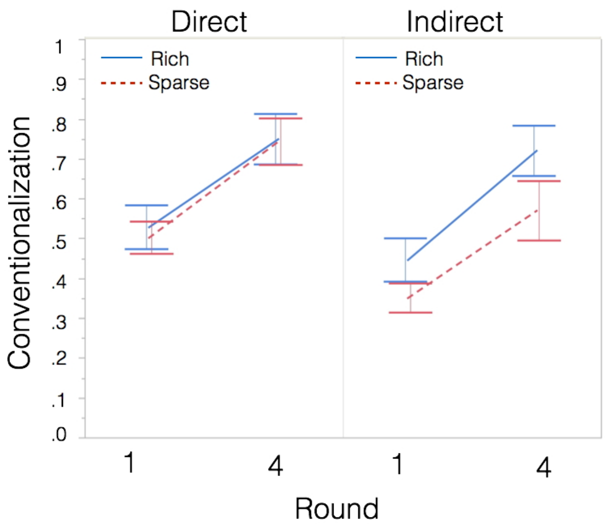

Figure 1. Conventionalization at the beginning (Round 1) and end (Round 4) of the experiment. Direct conventionalization (left) measures similarity between members of a given interacting pair (A & B, A & C, A & D). Indirect conventionalization (right) measures similarity among utterances that were produced by different participants (B, C, D) to a common interlocutor (A). Error bars represent 95% confidence intervals.

fully-connected network shows more conventionalization than the sparsely connected network [$F(1,12) = 17.82$, $p < .01$]. This is the central result.

2.3. *Number of interactions*

The advantage for the full over sparse network was observed even in Round 1, and there was no different in the rate of increase from Round 1 to Round 4 [$F(1,12) = 2.48$, $p = .14$]. This was unexpected. However, one consequence of the network structure manipulation was that the fully connected network entailed, by design, *twice as many communicative interactions per round as the sparsely connected network* (see Table 1). To address whether this simple property could in fact be the mechanism underlying the group difference, we compare the end-state of the sparsely connected network (Round 4, after 12 interactions) against the midpoint of the richly connected network (Round 2, after 12 interactions). There is no difference in conventionalization between the two groups once number of interactions is controlled [$F(1,12) = .21$, $p = .66$], suggesting that the greater number of interactions in the richly connected network may be responsible for the bulk of the effect. To reiterate, it appears that the fully- and sparsely-connected networks achieve similar levels of group-wide conventionalization after comparable numbers of interactions.

3. Discussion

The emergence of referring expressions is an important milestone in the development of any language-like system. Early on, different members of a community might use expressions that differ in both their semantic content and their form; however, conventionalization is likely to happen over time. We used actual human behavior in a natural communicative modality to show that the structure of the communicative network can influence the process of semantic conventionalization. Richly-connected communities (where agents communicated with all other agents using the emerging system) were more fully conventionalized (at least semantically) than sparsely-connected communities (where one agent used the emerging system with all others, but those others did not communicate amongst themselves in the emerging system).

In our data, this effect appears to be driven primarily by an epiphenomenon of the network structure manipulation: the fact that, all else being equal, a richly connected network will involve more communicative interactions than a sparsely connected network. This finding contrasts with some previous computational and behavioral work: Judd et al. (2010) found that humans reached consensus on an arbitrary choice of a color fastest when connected in a richly connected network, an effect unexplained by number of interactions. Similarly, simulations by Richie et al. (2014) showed an effect of network structure even after controlling for the number of interactions.

These contrasting effects reveal the subtleties of network effects, and illustrate the insufficiency of our intuitions about what networks might facilitate collective behavior. This point is even further driven home by consideration of simulation results by Gong, Baronchelli, Puglisi, and Loreto (2012), also a simulation study into the effects of social network structure on the emergence of language. In their model, agents must carve a perceptual continuum (color) into perceptual categories, and then agree upon labels that refer to one or more perceptual categories. In contrast to our model, where the communicating agents know the referent (e.g., avocado) but adjust the probabilities of corresponding gestures, Gong et al.'s agents must infer the referent through perception, and they found that sparsely- and richly-connected networks offered comparable convergence properties. Clearly, even though all the foregoing work involves conventionalization or consensus-building of some kind, network effects seem to be highly sensitive to assumptions/simplifications made by models, and/or the precise nature of the task set before participants.

The present results generate a hypothesis about why some naturally-emerging sign languages appear to conventionalize their referring expressions (at the semantic level) more quickly than others: they may simply be communicating more. However, this remains to be empirically verified, and other alternatives remain to be explored. For example, it could also be that conventionalization within a local community (e.g. a single family) could impede larger-scale conventionalization among members of different families, schools, villages, etc. Because the present study tested only a single community in which all members interacted, the results cannot yet speak to this issue.

3.1. *Future directions*

To date, we have focused on characterizing conventionalization at the semantic level. Future work will explore conventionalization at the level of form. We are particularly interested in testing the hypothesis that the emergence of syntax may not require that referring expressions be conventionalized, but the emergence of phonology might. This would be consistent with reports of emerging sign languages in which syntactic structure is evident, but phonological structure is not. In addition, simulating the impact of local conventionalization on global conventionalization is a major target for both computational and behavioral experiments.

References

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

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

Judd, S., Kearns, M., & Vorobeychik, Y. (2010). Behavioral dynamics and influence in networked coloring and consensus. *PNAS, 107*(34), 14978–14982.

Meir, I., Sandler, W., Padden, C., & Aronoff, M. (2010). Emerging sign languages. *Oxford Handbook of Deaf Studies, Language, and Education*, *2*, 267-280.

Morgan, H. (2015). When does a word emerge? Lexical and phonological variation in a young sign language. *San Diego Linguistic Papers, 5,* 2-17.

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

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.

Steels, L. (2011). Modeling the cultural evolution of language. *Physics of Life Reviews*, *8*(4), 339-356.