Journal of Language Evolution, 2025, **10**, 1–13
https://doi.org/10.1093/jole/lzaf003
Advance access publication 25 November 2025
Research Article

OXFORD

Discrete and systematic communication in a continuous signal-meaning space

Alicia M. Chen[1,*,‡], Matthias Hofer[1,‡], Moshe Poliak[1], Roger Levy[1], Noga Zaslavsky[2]

[1]Department of Brain and Cognitive Sciences, MIT, 43 Vassar St., Cambridge, MA 02139, United States
[2]Department of Psychology, NYU, 6 Washington Place, New York, NY 10003, United States

[‡]A.M.C. and M.H. contributed equally to this work.

*Corresponding author. Department of Brain and Cognitive Sciences, MIT, 43 Vassar St., Cambridge, MA 02139, United States.
E-mail: aliciach@mit.edu

Associate Editor: Kenny Smith

Human spoken language uses a continuous stream of acoustic signals to communicate about continuous features of the world, by using discrete forms—words—that segment the world into categories. Here we investigate how *discreteness* (the segmentation of a continuous signal space into discrete forms) and *systematicity* (the consistent alignment of these forms with what they refer to in the world) can emerge under communicative pressure. In an exploratory study, participants were paired with one another and played a game in which they varied the pitch of auditory signals to communicate about a continuous color space, generalizing from a small, shared set of signal-color pairings. The emergent systems exhibited both discreteness and systematicity, but only systematicity robustly predicted successful communication. These findings offer insight into the cognitive strategies that could support the creation and evolution of language, highlighting how pressures for effective communication can shape continuous signal spaces into structured, learnable systems.

Keywords: discreteness; systematicity; communication game; color categories.

One of the most intriguing aspects of human spoken language is that it exhibits both discrete and continuous properties. While language unfolds in a continuous medium—spoken words are transmitted as streams of acoustic signals that vary in pitch, duration, and intensity—listeners perceive these signals categorically (Liberman et al. 1967; Harnad 2003), allowing for clear distinctions between words like ‘bark’ and ‘park’ even though the underlying acoustic features are continuous. Similarly, while some conceptual spaces underlying word meanings—such as the perceptual continuum for color—are continuous, words such as ‘red’, ‘orange’, and ‘yellow’ correspond to concepts that carve out discrete categories within these spaces (Rosch 1973; Carey 2009). These categories are not inherently predetermined (Berlin and Kay 1969), yet they allow for clear categorical distinctions that facilitate the infinite productivity that defines human language.

Moreover, the organization of these discrete units of sound and meaning is not arbitrary. Linguistic forms and meanings are often systematically aligned in the form of word order (e.g., we interpret the distinct meanings of ‘the cat chased the dog’ and ‘the dog chased the cat’; Sandler et al. 2005; Schouwstra and de Swart 2014). And although relationships between form and meaning at the word level have traditionally been viewed as arbitrary (Hockett 1960), recent work has highlighted various ways in which systematicity is present even at this level, where sound patterns may carry meaning beyond arbitrary conventions (Monaghan et al. 2014; Dingemanse et al. 2015; Blasi et al. 2016; Pimentel et al. 2019). For example, similar-sounding words like ‘glow’, ‘gleam’, and ‘glimmer’, have phonological features that cue their related meanings (Bergen 2004; Sidhu 2025). Likewise, a color naming system that uses the word ‘grue’ to label intermediate green-blue hues, would demonstrate systematicity.

Thus, in a continuous meaning space, using a continuous physical medium, discreteness and systematicity critically work together to shape human language.

Discreteness segments an infinite set of sounds and meanings into a finite set of categories that we communicate about, thus providing the foundational building blocks for communication. And systematicity organizes these building blocks in ways that render form-meaning relationships highly transparent and predictable, benefiting acquisition and learnability (Dingemanse et al. 2015; Raviv et al. 2021), processing and use (Monaghan et al. 2012; Nölle et al. 2018), as well as transmissibility and evolvability (Kirby et al. 2008; Xu et al. 2013; Raviv et al. 2019; Morin et al. 2022; Motamedi et al. 2022). In other words, given constraints on information processing and learning (e.g., Gibson et al. 2019), these strategies simplify and structure the form and meaning space, so that we can successfully communicate with each other. In the current work, we focus on the cognitive strategies underlying this link: how discreteness and systematicity might emerge to facilitate communicative goals.

Prior research has shown that both discreteness and systematicity can independently emerge in artificial languages, but primarily in settings without real-time communication and without the continuous signal and meaning spaces that are reflective of human language. These studies often use already-discretized signaling spaces paired with continuous meaning spaces (e.g., Xu et al. 2013; Carr et al. 2017), or continuous signaling spaces paired with with already-discretized meaning spaces (e.g., De Boer and Verhoef 2012; Verhoef et al. 2015, 2016). Nonetheless, systematicity has been observed to emerge in real-time communication games involving continuous signal spaces and discretized meaning spaces (Theisen et al. 2010; Verhoef et al. 2016). Additionally, computer simulations have shown that both discreteness (Lieck and Rohrmeier 2021) and systematicity (Zuidema and Westermann 2003) can emerge as optimal solutions to communication problems, but these studies have not explored how discreteness and systematicity interact under communicative pressure in real-time communication settings.

To directly study how these strategies emerge to support communication, we designed an online communication game (see Müller and Raviv 2021, for a review on the use of communication games for studying language evolution) using a minimal artificial setting where *both* the signal and meaning space are continuous, and where we could apply standard analysis techniques to investigate the emergence of discreteness and systematicity. More specifically, participants used whistled signals (Verhoef 2012), to communicate about colors. They had to create a communication system with their partner that generalized a small amount of 'common ground' (a set of arbitrary signal-color pairings) to a larger set of colors. Thus, in our setting, both discreteness and systematicity are initially absent, but could emerge under communicative pressure. We conducted a set of exploratory analyses testing whether discreteness and systematicity emerged in the first place, as well as which of these strategies helped people communicate successfully in the game.

1. Methods

Our experiment consisted of a *learning phase* immediately followed by a *communication phase* (see Fig. 1). In the learning phase, each participant learned the same set of five signal-color mappings shown in Fig. 1a, initializing their 'common ground'. In the communication phase, participants were then paired up and asked to communicate forty colors to each other, extrapolating from the learned signals to communicate about colors they had not encountered before.

1.1 Materials

1.1.1 Colors

The forty colors that participants had to communicate are shown in the color wheel of Fig. 1b. They are a slice from the World Color Survey's (WCS) standard color naming grid (Berlin and Kay 1969), with a fixed brightness level (row F in the WCS grid). This choice of colors captures the hue dimension while maintaining a reasonable experiment length. Of these, the five colors utilized in the learning phase (Fig. 1a) were randomly chosen to be approximately equidistant, with slight random perturbations of $\pm 1 - 3$ color chips away from the initial selections to introduce perceptual irregularity.

1.1.2 Signals

The signals were produced by an on-screen slide whistle interface similar to Hofer and Levy (2019), represented as sequences of pitch over time (Fig. 1a). In our experiment, this whistle was visually represented as the trunk of an alien creature, which participants controlled using the mouse and space bar. Stretching the creature's trunk decreased the pitch, while shortening it increased the pitch (see https://osf.io/ynbp6/ for a video sample). To obtain the five initialization signals (Fig. 1a), we selected signals with a variety of structured perceptual features, from a larger set of signals previously evaluated in a norming study conducted by Hofer and Levy (2019) in which subjects were asked to rate signals according to their relative complexity and perceptual similarity.

1.2 Procedure

1.2.1 Learning phase

During the learning phase, each participant learned the same five signal-color mappings (Fig. 1a). First, participants

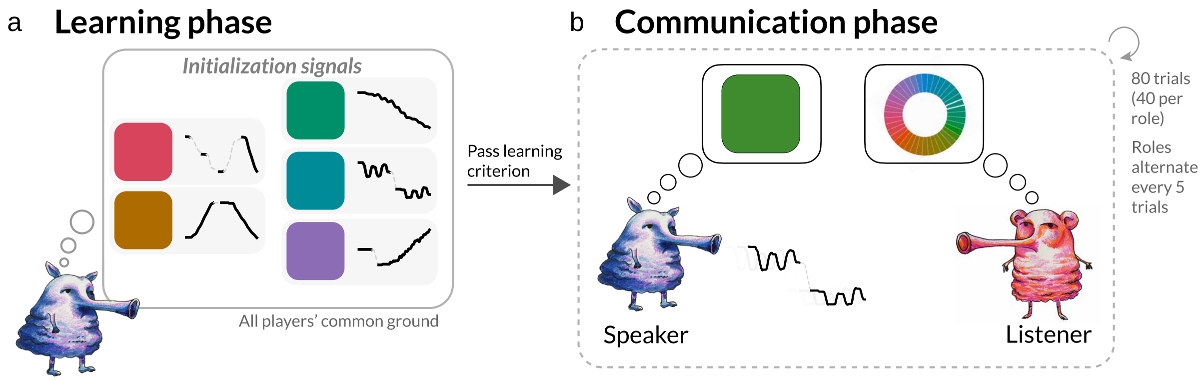

Fig. 1. Experimental setup. In the learning phase (a), participants learn five initialization signal-color pairings (signals shown are a visualization of pitch over time). In the communication phase (b), participants are assigned speaker and listener roles, and have to extrapolate their learned signals to communicate about a total of forty colors.

heard the signal for each color and were instructed to use their own slide whistle to reproduce the signal. This was done five times for each signal-color pairing. Then, we tested how well participants remembered the signal-color mappings: For each pairing, the signal was played while participants saw the target color and a distractor color (pseudo-randomly selected from the remaining four colors), and had to guess the color corresponding to the signal. Participants passed if they correctly guessed four of the five pairings. No feedback was given for this phase of the task. Finally, to measure how well participants actually remembered each signal, we asked participants to reproduce the signals for each of the five colors from memory.

1.2.2 Communication phase

After passing the learning phase, participants were grouped into pairs and randomly assigned speaker and listener roles that alternated every five trials. In each trial, the speaker produced a whistled signal to convey a designated target color to the listener. The listener then indicated their guess on a color wheel containing forty colors (see Fig. 1). Once the trial ended, both participants saw the target and the guessed color, and bonus payments were given based on how accurate the listener's guess was in terms of perceptual distance in color space (see Analysis Techniques section below for specifics). Participants were given blocked feedback about how close their guesses were, on average, across the last 10 trials. Each participant had forty trials as speaker and forty as listener, with the order of the colors and the color wheel orientation randomized between participants. See Fig. A.1 for screenshots of the task.

1.3 Participants

We recruited sixty-five native English speakers from the USA or Canada with normal or corrected-to-normal vision, from the crowdsourcing platform Prolific. Participants were prescreened using a color sorting task (Foutch et al. 2011) and a short audio task (Woods et al. 2017) to verify that they have normal color vision and that they were using headphones. Of these, fifty-four participants passed the learning phase and advanced to the communication phase, with fifty participants successfully completing the communication phase, resulting in twenty-five full games. Compensation was variable based on progress in the study, but averaged 15 per hour. Participants received an additional bonus of up to approximately 10% of their earnings contingent on their performance in the learning and communication tasks. Participants gave informed consent, and all procedures were approved by the MIT Committee for the Use of Humans as Experimental Subjects.

1.4 Analysis techniques

1.4.1 Measuring similarity between signals and referents

To examine participants' communication systems, we analyze our data using measures of similarity, classically used to study people's mental representations in various domains including color (Shepard 1987; Shepard and Cooper 1992).

1.4.1.1 *Similarity between signals*

The signals participants produced in the experiment are time series data represented as pitch over time. To determine similarity between signals, we used Dynamic Time Warping (DTW; Sakoe and Chiba 1978; Berndt and Clifford 1994), following its use in prior work to analyze continuous signals in cultural transmission experiments (e.g., Verhoef 2012). DTW is a technique used to measure the similarity between two sequences that may vary in speed or duration. Unlike standard distance measures that compare signals point-by-point at corresponding time intervals, DTW flexibly aligns sequences by allowing the temporal axis to be stretched or compressed. This alignment process matches elements of the two sequences based on their overall shapes, effectively minimizing differences in timing. Thus, even if one signal spans two seconds and another spans five seconds, they can still be considered highly similar if their overall shapes closely match (refer to Fig. 1 for example signals). DTW is particularly useful in analyzing human-generated signals such as speech or gesture, where timing often varies naturally but the overall patterns remain consistent and are more informative for their meaning.

1.4.1.2 *Similarity between referents*

We used Euclidean distance in CIELUV color space to quantify similarity between colors. The CIELUV color space represents colors using three numerical values: L*, u*, v* (chromaticity coordinates). It is specifically designed so that the Euclidean distance between two points corresponds to the perceptual distance perceived by human observers. In other words, two colors that are numerically closer in this space appear more similar to human vision than colors that are numerically farther apart (Schanda 2007).

1.4.2 Visualizing signal similarities in a lower-dimensional space

To qualitatively visualize and compare participants' signal repertoires, we calculated signal similarities for all signal pairs across participants using DTW, allowing us to embed and visualize the data in a lower-dimensional space using multidimensional scaling (MDS; Kruskal 1964). MDS is a statistical technique that represents similarities between items as distances in a lower-dimensional space, preserving the relative relationships observed in the original, higher-dimensional data. It has classically been used to visualize human similarity judgments, such as perceived similarities among objects, concepts, or perceptual stimuli (Shepard and Cooper 1992). In our study, MDS allows us to project information about the signals into a space that we can visually interpret, as well as use for further calculations.

1.4.3 Discreteness and systematicity

We next turn to how we measured discreteness and systematicity—the critical features of participants' communication systems that we were interested in.

The conceptual examples in Fig. 2 display different combinations of discreteness and systematicity that participants' communication systems can exhibit. In this visualization, each marker represents an individual signal (colored by its corresponding referent), positioned along a one-dimensional signal-similarity space. The distance between markers represents perceptual differences between signals, such that signals closer together are perceived as more similar. This simplified representation corresponds to a 1D projection computed using MDS, on DTW similarities.

1.4.3.1 *Discreteness*

Human language segments the continuous spectrum of human experience into discrete categories marked by categorical boundaries, even in continuous perceptual spaces like colors, motion, spatial relationships, and time. In our conceptual examples, a signaling system exhibits high *discreteness* (Fig. 2, rows 1 and 2) if signals clearly cluster into distinct groups. Such clusters are analogous to linguistic words—discrete entities that map onto sets or ranges of referents.

To quantitatively measure discreteness, we operationalized discreteness in two complementary ways, each capturing distinct intuitions about what it means for a system to be discrete.[1] First, we calculated the Hopkins statistic (Lawson and Jurs 1990), computed on MDS embeddings. The Hopkins statistic quantifies how clustered a system is, by comparing distances between points and their nearest neighbors in the dataset against distances from points in a uniformly distributed null. Values range from 0 (highly regular distribution with minimal clustering) to 1 (highly clustered), with 0.5 indicating a random uniform distribution. This metric captures the overall tendency for signals to form distinct clusters rather than being evenly spread out from each other.

We also measured discreteness using the number of clusters detected by Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN; Campello et al. 2013). HDBSCAN is a nonparametric clustering algorithm that identifies clusters based on local density, requiring only one major free parameter (minimum points per cluster). It is robust to varying cluster shapes and densities, and operates directly on the DTW signal similarities, making it particularly suitable for detecting natural groupings within signal spaces. HDBSCAN identifies distinct clusters in regions where signals densely concentrate, and returns a single cluster if the signals are uniformly distributed. This metric captures a more indirect notion of discreteness: the notion that discrete systems consist of multiple clearly defined groups of similar signals (analogous to the number of words in a language).[2]

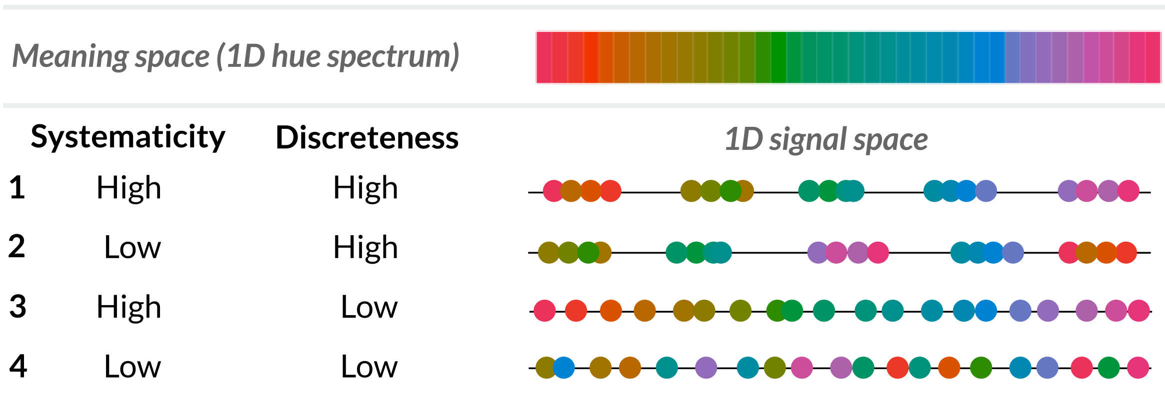

Fig. 2. The discreteness-systematicity hypothesis space. Each signal, associated with a referent (its color), is represented by a point in a one-dimensional signal-similarity space.

1.4.3.2 Systematicity

A signaling system has high *systematicity* (Fig. 2, rows 1 and 3) if signal forms provide reliable cues to their meanings. In the context of our experiment, the existence of systematicity implies a consistent relationship between the similarity of signals and the similarity of their associated colors. For example, signals representing similar colors (such as blue and green) should sound more alike than signals representing more distinct colors (such as blue and yellow).

To quantify systematicity, we computed the distance correlation between pairwise DTW distances (signal similarities) and CIELUV distances (color similarities). Distance correlation is a measure that captures both linear and nonlinear relationships between two sets of distances (Székely et al. 2007), and is analogous to the way systematicity has been calculated in previous work (e.g., Mantel 1967; Kirby et al. 2008; Verhoef et al. 2015; Atkinson et al. 2019). Distance correlation ranges from 0, indicating no relationship (independence), to 1, indicating a perfect deterministic correspondence between signal similarity and referent similarity.

2. The discreteness-systematicity hypothesis space

What could participants' extrapolated communication systems look like? The most straightforward strategy is that participants simply reproduce their signals into perceptually similar clusters around the nonsystematic initialization signal-color pairings (Fig. 2 row 2; low systematicity high discreteness). Alternatively, participants could come up with new, systematic signals to refer to each cluster (Fig. 2 row 1; high systematicity and high discreteness)—this strategy may be more challenging because participants have to come up with new agreed-upon signal-meaning pairs, but the increased systematicity could facilitate communication because of increased expressivity.[3] Finally, strategies that could theoretically enable participants to precisely communicate all of the colors are communication systems where each signal corresponds to a distinct meaning (i.e. low discreteness). Such a system could either exhibit high (Fig. 2 row 3) or low (Fig. 2 row 4) systematicity.

From this, we can see that systematicity and discreteness can interact with each other in nuanced ways. A maximally systematic language would be highly expressive, capable of communicating fine-grained differences in meaning, but could impose cognitive pressures on learning and memory. By contrast, a discrete system might be less expressive but more conducive to learning and coordination. Simply reproducing initialization signals for perceptually similar clusters could facilitate communication by reducing miscoordination risks, while forming new systematic communication systems might enhance learnability and informativity.

3. Results

3.1 Performance in the learning and communication phases

We started by evaluating how well participants were able to learn the signals and generalize their learned signals to communicate with their partners, in the first place. To this end, we assessed the performance in the learning and communication phase. For the learning phase, we measured performance by how accurately people reproduced the five signals at the end of the

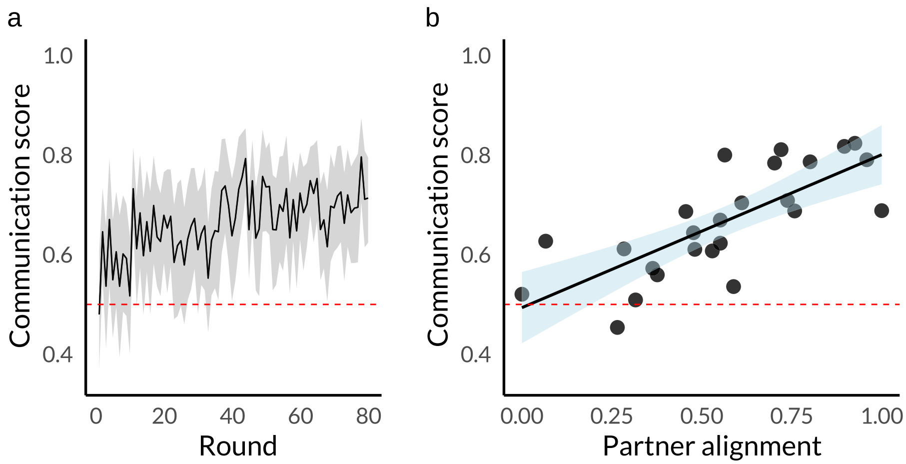

Fig. 3. (a) Communication score (black curve) as a function of the number of online communication rounds. With time, participants learn to communicate successfully. (b) Communication score (black curve) as a function of the alignment between the dyad's signal repertoires. Each point represents one game, and the *y*-axis corresponds to the mean of the two players' communication score within a game. As expected, partners with more aligned repertoires communicated better. In both plots, shaded regions correspond to 95% confidence intervals and dashed red lines correspond to baseline communication scores (i.e. random guesses).

learning phase. Specifically, we calculated the average DTW distance between participants' reproductions and the original initialization signals. This average was then normalized across participants to obtain a *learning score*, with 0 indicating minimal learning and 1 indicating the most learning.

We assessed the communication phase performance based on how accurately speakers guided their partner to the target color. For each trial, communication accuracy was determined using CIELUV distances between the target color and the guessed color, normalized between 0 (identical colors) and 1 (the most perceptually distinct pairs among the forty experimental color patches).

Intuitively, we expect participants that are better learners to do better in the communication phase because they would be able to, on average, more accurately recall their partners' signal-color associations and communicative tendencies. This was indeed the case (linear regression predicting communication score from learning score; $b = .241$, $t(45) = 3.022$, $p = .004$; Fig. A.2).

To further assess how communication performance evolves as a function of interaction time, we used a linear mixed-effects model, with random intercepts for game and participant, to predict communication score from round (each participant communicated about the forty colors, so there were eighty rounds in total). Participants performed above chance overall (Fig. 3a; M = 0.665, SE = 0.022, CI: [0.62, 0.71], $t(24) = 7.699$, $p < 0.001$). Performance increased across rounds ($b = 0.002$, $t(1974) = 6.654$, $p < 0.001$). This suggests that participants were able to successfully establish communicative conventions that allowed them to go beyond the systems with which they started, consistent with findings in related work on convention formation in reference games (e.g., Hawkins et al. 2023).

3.2 Communication strategies

Having established that participants successfully learn to communicate in our experimental setting, we next turn to assessing what specific strategies enabled successful communication.

3.2.1 Representing and visualizing signals

We computed the MDS embeddings using the pairwise DTW signal similarity calculations. The quality of MDS solutions is typically evaluated using stress values, which quantify discrepancies between the original similarity distances and the distances in the lower-dimensional embedding. See Fig. A.3 for MDS stress values as a function of the number of dimensions retained.

Figure 4b depicts an MDS representation of all signals across all games, including the five initialization signals from the learning phase. Each signal is colored by its associated referent color, and the intialization signals are marked in boxes. Signals generally cluster around the five initially learned signal-color associations. More surprising, however, is the emergent structure between the produced signals. Some signals migrated to form a systematic hue gradient: For

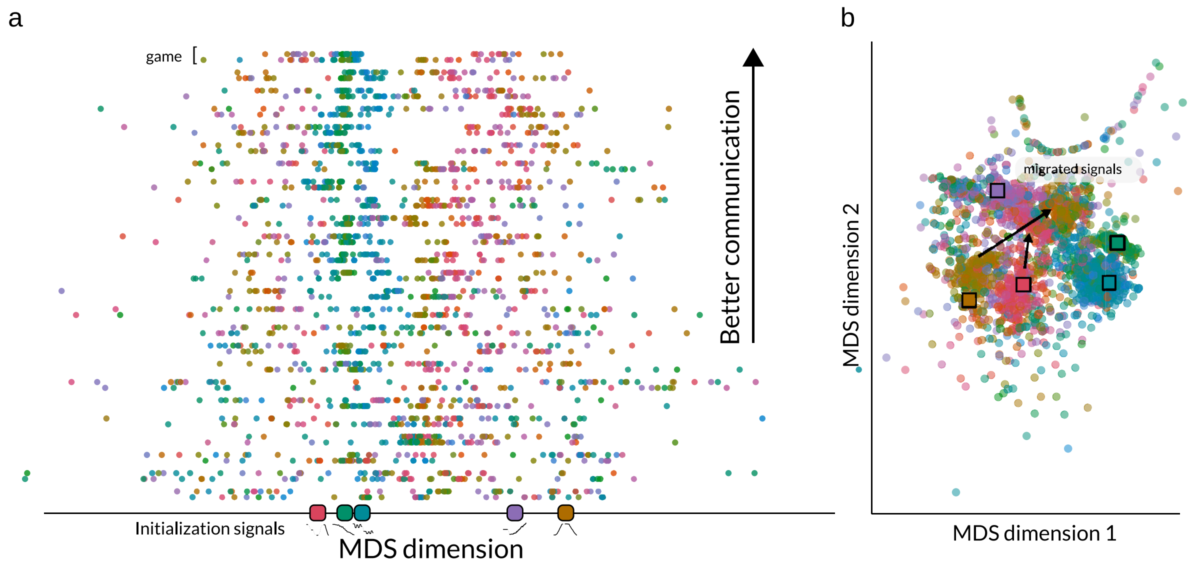

Fig. 4. Qualitative assessment of the use of systematicity. (a) A one-dimensional MDS embedding of the emergent signaling systems. Each pair of lines corresponds to the two partners in a game. Each dot corresponds to a signal, and is colored by its target color. The games are sorted by their average communication score. (b) A two-dimensional MDS visualization of all emergent signals across all dyads. As in (a), each dot corresponds to a signal and is colored by its target color. Black squares correspond to the initialization signals.

example, yellow and pink signals migrated to be between the initial purple and green signals.

To assess whether similar systematic patterns appear across individual participants, we visualized participants' signals in a one-dimensional embedding space, ordered by communication score (Fig. 4a). These embeddings indicate that this systematic structure persists at the individual level: Participants produced intermediate signals to interpolate between the learned signal-color pairings, in a systematic way. Participants who were the least successful at communicating displayed low systematicity and discreteness, while participants who communicated better seemed to display higher discreteness as well as emergent systematic structure. The raw signals for each of the games are visualized at https://osf.io/mwnfy. Next, we assess these qualitative observations quantitatively.

3.2.2 Partner alignment

What properties could drive successful communication? The first dimension of interest we consider is how much the two players' signals align, within a game. Successful communication is often characterized by the emergence of shared conventions (Lewis 1969; Hawkins et al. 2019), implying that successful dyads in our task would produce similar signals to refer to the same color. To quantify this, for each game we calculated the similarity of the two partners' repertoires: For each color, we calculated the similarity between the two partners' signals, and averaged this measure across all referents to obtain a measure of *partner alignment* for each game (normalized to be between 0 and 1, where 1 indicates the most alignment). The degree to which partners' signals were aligned, was correlated with how well they communicated with each other (alignment vs. within-game average communication score; Pearson's $r(23) = 0.752$, $t(23) = 5.465$, $p < 0.001$) (Fig. 3b).

We tested how much this effect is explained by how well individual participants are able to remember signal-color pairings in the first place. In a linear regression predicting communication score from partner alignment and learning score, we found a main effect of alignment but not of learning score (alignment $b = 0.288$, $t(44) = 6.202$, $p < 0.001$; learning score $b = 0.012$, $t(44) = 1.144$, $p = 0.259$), suggesting that memory for specific signals does not itself drive participants' ability to align their signals to each other and do better in the task.

3.2.3 Discreteness

Next, we investigate what properties of participants' emergent communication systems *did* help them communicate better in the task. We started by calculating the Hopkins statistic computed on three-dimensional MDS embeddings.[4] In three-dimensional embedding space, participants' systems tended to be discrete; M(SD) discreteness = 0.75(0.06). We then measured

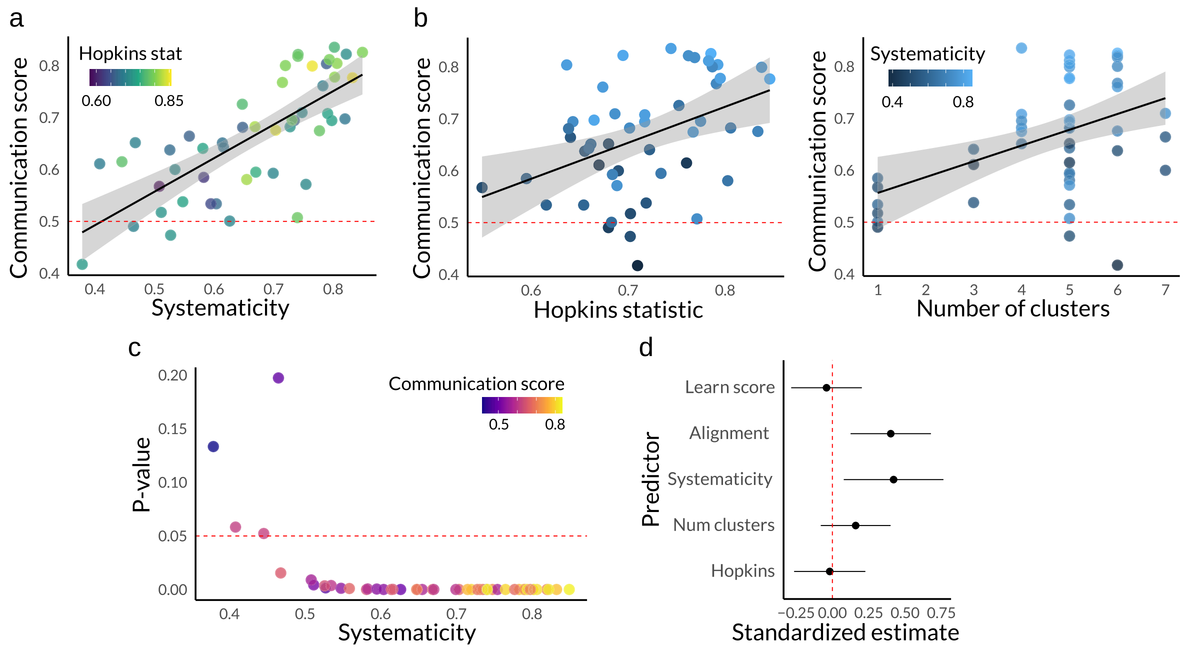

Fig. 5. Systematic (a) and discrete (b) systems were better for communication. Most participants produced systematic communication systems (c), and partner alignment and systematicity predicted communication score (d). Error regions are 95% confidence intervals.

discreteness using number of clusters found using HDBSCAN. In our data, the number of clusters detected (HDBSCAN with a minimum of two points per cluster) ranged from one cluster to seven clusters. The most common number of clusters found was five (in twenty-three out of fifty participants); we visually validated that in these systems, participants typically clustered their signals around the five initialization signals.

The two discreteness measures were correlated with each other (Pearson's $r = 0.490$, $t(48) = 3.892$, $p < 0.001$). They were also both positively correlated with communication score (Hopkins statistic Pearson's $r = 0.420$, $t(48) = 3.209$, $p = 0.002$; number of clusters Pearson's $r = 0.499$, $t(48) = 3.985$, $p < 0.001$; Fig. 5b).

3.2.4 Systematicity

For each participant, we measured systematicity by calculating the distance correlation between pairs of signals and referents. Participants' signaling systems were generally systematic (M(SD) = 0.67(0.12)). For each participant, we tested whether the calculated systematicity was above and beyond what would have been produced by chance, by permuting the assignment of signals to referents (10000 permutations per participant). The p-values for forty-six of the fifty participants were below 0.05, with the majority of p-values below 0.001 (Fig. 5c) The initialization signals were not systematic ($p = 0.171$). Across participants, systematicity was positively correlated with communication score (Pearson's $r = 0.733$, $t(48) = 7.457$, $p < 0.001$; Fig. 5a).

3.2.4.1 *Systematicity in the presence of discreteness*

Our measure of systematicity captures but does not directly distinguish between different ways systematicity can emerge: Systematicity, operationalized on the signal level, can also act on the cluster level (i.e. similarity between clusters is informative for their meaning) if clusters exist. Moreover, a maximally systematic system cannot be discrete, but a discrete system where the clusters are not organized in a systematic way (i.e. Fig. 2, row 2) would still display mild systematicity because similar signals are organized close together into clusters.

We tested how much our clustering method can distinguish between whether systematicity emerged within- or between-clusters. For each participant, we calculated within-cluster systematicity for each cluster, and between-cluster systematicity if more than three clusters were detected using HDBSCAN with `min_cluster_size = 2`, by calculating systematicity using the cluster medoids (the point in the cluster with the minimal distance to the other points in the cluster) and their corresponding referents. Most participants did not exhibit statistically significant amounts of between-cluster systematicity (forty-two out of forty-four participants for which three or more clusters were detected). We also did not find evidence of within-cluster systematicity in most of the clusters produced across participants (twenty-seven clusters across twenty-three participants, out of 229 total clusters across participants). For `min_cluster_size = 2`, between-cluster systematicity was weakly correlated with general systematicity (Pearson's $r = 0.305$, $t(42) = 2.075, p = 0.044$), but this was not robust across different values (2–7) of the `min_cluster_size` parameter. Between-cluster systematicity was not correlated with communication score (Pearson's $r = 0.165$, $t(42) = 1.085$, $p = 0.284$); this was also the case for the other values (3–7) of the `min_cluster_size` parameter that we tested (but note the difficulty of comparing the amount of between-cluster systematicity between systems with different numbers of clusters). Thus, with our data and methods, we cannot make conclusions about whether either between- or within-cluster systematicity emerged to support successful communication. Although participants did seem to form clustered systems, systematicity seems to instead primarily be used to form a continuous gradient to interpolate between other clusters (Fig. 4a).

3.3 Systematicity and partner alignment predict communication score

Finally, we conducted a linear regression predicting communication score from all of the above (scaled) predictors: learning score, partner alignment, systematicity, number of clusters, and Hopkins statistic. There was no problematic multicollinearity (variance inflation factors 1.64–3.08). Partner alignment ($b = 0.402$, $t(41) = 2.934$, $p = 0.005$) and systematicity ($b = 0.421$, $t(41) = 2.480$, $p = 0.017$) were significant predictors of communication score (Fig. 5d), while the other predictors, including both discreteness metrics, were not (learning score $b = -0.040$, $t(41) = -0.335$, $p = 0.740$, Hopkins stat $b = -0.018$, $t(41) = -0.150$, $p = 0.881$, n-clusters $b = 0.160$, $t(41) = 1.347$, $p = 0.185$).

3.4 Qualitative analysis of individual strategies

Some participants self-reported discovering novel techniques for combining signals to communicate. Several participants showed signs of compositional strategies that combine two already-existing signals to refer to in-between colors (e.g., participants combined signals for red and yellow to refer to orange). Other participants reported modifying continuous aspects of signals, such as their lengths, to communicate to their partner the distance between target color was from the color that the signal was 'prototypically' associated with.

These results have direct bearing on debates about the emergence of combinatorial signaling. For instance, the first strategy described is essentially the 'synthetic route' to combinatoriality outlined in Zuidema and De Boer (2018), where combinatorial structure emerges by building up complexity through the concatenation of entire signals. Although such features of the data cannot easily be captured with the signal-processing methods employed here (but see Hofer et al. 2021, for work on more expressive modeling techniques), investigating the emergence of signal-internal (combinatorial) structure or meaning-dependent (compositional) structure in the context of this experimental paradigm is an important next step.

4. Discussion

We investigated how discreteness and systematicity interact to support communication in continuous signal-meaning spaces. In our study, participants learned a small set of highly structured and nonsystematic signals to refer to colors. They then were paired with another participant and asked to generalize these associations to a larger space of colors in a reference game. Participants produced novel signals with their partners, to interpolate between the initialization signals in a systematic way. Discreteness and systematicity were both correlated with communication performance, and systematicity specifically predicted successful communication.

Why might discreteness not be a predictor? We might speculate that the compression of a communication system into discrete signals is important for satisfying constraints other than successful communication, such as learnability or memory constraints. Discreteness may help participants generalize the learned signals to a larger set of signals, but may not itself drive communication. Additionally, the way that systematicity is calculated also captures some aspects of discreteness: Systems that are clustered would be detected as moderately systematic if each cluster refers to a similar set of referents.

One limitation of our study is that we cannot assess whether discreteness actually emerged from scratch. Instead of assessing how participants establish an entirely new communication system, we chose an arbitrary set of signal-color pairings to seed as 'common ground'. These initialization signals might have biased participants toward discreteness (i.e. forming clusters around the initialization signal-color mappings). The reason for using initialization signals and a learning phase was to make emergent systems more comparable across participants, and maintain a reasonable task length given the difficulty of establishing a new communication system from scratch (especially when participants cannot rely on iconicity to bootstrap communicative conventions; Macuch Silva et al. 2020).

Our study shows how a continuous, one-dimensional meaning space, along with repeated dyadic communication, imposes cognitive pressures on continuous systems that lead to the use of discrete and systematic strategies. These results imply that continuous meaning spaces themselves may inherently impose learnability pressures and drive structural organization in communication. It is plausible that increasing the dimensionality of the color space, such as including saturation or brightness, could lead to additional pressures for structured signaling beyond what what we have observed (De Boer and Verhoef 2012; Little et al. 2017). Another possible source of communicative pressure is generational transmission, where one generation's signal-color mappings are passed on to the next (Kirby et al. 2015).

The minimal artificial setup in our study allows us to isolate the contribution of certain properties of communication systems in supporting successful communication. However, the continuous signal-meaning space we use opens the door for many further questions, relating to existing properties in natural language. For example, one question is whether and how participants' emergent communication systems are influenced by the color categories present in their native language (Xu et al. 2013). While we test native English speakers, who likely have cognitive biases shaped by English color categories, participants in our experiment did not cluster their generalized signals around categories that neatly align with English color terms. However, over generations, emergent systems could reflect categories in natural language and/or information-theoretic principles of color naming (Xu et al. 2013; Zaslavsky et al. 2018).

Finally, our current analyses focus on the similarities *between* signals and their referents, based on the complete repertoires that participants produced over the eighty rounds. These analyses are important for studying the relationship across signals, but do not consider qualitative or quantitative structure *within* signals themselves, or the temporal dynamics of social interaction. Previous work using continuous modalities, notably with whistled signals, have found the emergence of signal-internal structure in the form of combinatorial building blocks, which we did not explicitly look for here (Verhoef et al. 2015; Hofer and Levy 2019). Although participants did self-report using compositional strategies, using DTW as a similarity metric cannot capture emergent structure within signals. Compositional strategies may be even more necessary in a larger referent space that varies in additional dimensions (Little et al. 2017; Lieck and Rohrmeier 2021). Therefore, one direction for future research is exploring how compositional building blocks might emerge within sound/phonetic space, that complement the emergence of the kinds of systems we see in our experiment. Furthermore, investigating the temporal dimensions of social interactions could shed light on how such conventions form and stabilize over time.

In sum, we studied people's communicative strategies when they used continuous signals to communicate about a continuous referent space. The partners that were better at this were able to change their signals from the initialization signals, building systematic communication systems that allowed them to communicate better over rounds. Studying the emergence of and change in communication in a continuous signal-meaning space may shed light on the cognitive processes that allowed humans to start using language.

Notes

1. Previous work has measured discreteness in various ways, including human rater judgments (Goldstein 2003; Sandler et al. 2011; Grice et al. 2017) and categorical perception tasks in existing languages (Newport 1982; Gussenhoven 1999).
2. While this measure of discreteness is intuitive for small numbers of detected clusters—for example, a single cluster implies a uniform distribution, while a few clusters indicate discreteness—it becomes less interpretable at higher values. It's unclear, for instance, whether detecting seven clusters meaningfully reflects greater discreteness than detecting six.
3. Note that we only look at the subset of discrete systems that are convex; i.e. signals cluster but all signals within a cluster refer to perceptually adjacent colors.
4. The number of MDS dimensions to retain was decided using the elbow method; see Fig. A.3

Acknowledgements

We thank our two anonymous reviewers, members of MIT's Computational Psycholinguistics Lab and Saxelab, Nathaniel Imel, as well as the anonymous reviewers and audience at CogSci 2022 and Evolang XV, for helpful discussions and suggestions.

Funding

This work was supported by a grant from the Simons Foundation to the Simons Center for the Social Brain at MIT.

Data availability

The data and code underlying this article are available at https://github.com/aliciamchen/color–whistles, and also deposited on OSF at https://doi.org/10.17605/OSF.IO/NQY6H.

Appendix

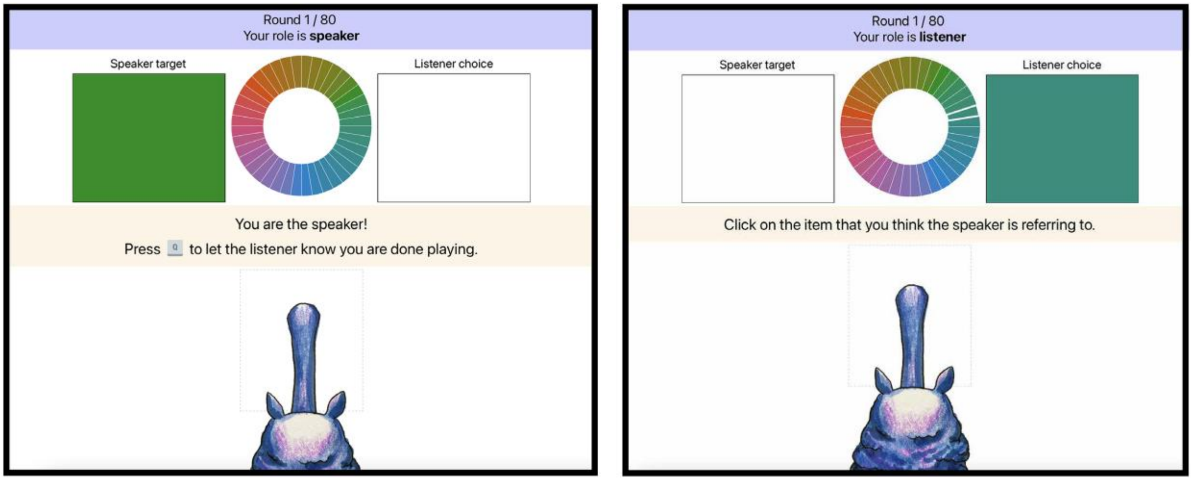

Fig. A.1. Screenshots of task. The speaker (left) moves the trunk of the creature to produce a signal to refer to a target color. The listener (right) hears the the signal (while seeing the trunk move on the screen), and selects what they think the target color is, by clicking on the color wheel.

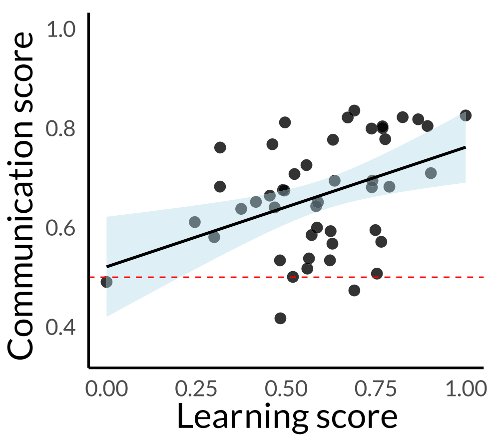

Fig. A.2. Communication score as a function of learning score. Each point represents one participant, and the *y*-axis corresponds to their partner's performance in the communication phase.

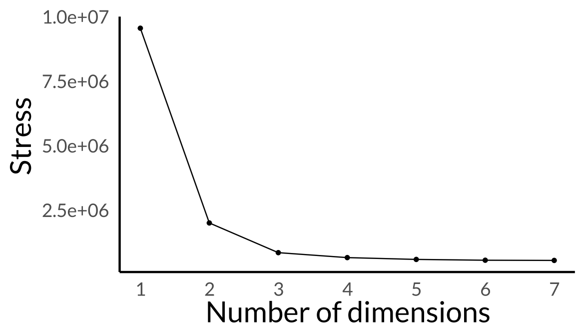

Fig. A.3. MDS stress values plotted against number of dimensions. MDS projections were calculated using all of the pairwise distances for all the combinations of signals, across all participants.

References

Atkinson M., Mills G. J., and Smith K. (2019). 'Social Group Effects on the Emergence of Communicative Conventions and Language Complexity', *Journal of Language Evolution*, 4: 1–18. https://doi.org/10.1093/jole/lzy010

Bergen B. K. (2004). 'The Psychological Reality of Phonaesthemes', *Language*, 80: 290–311. https://doi.org/10.1353/lan.2004.0056

Berlin B., and Kay P. (1969). *Basic Color Terms: Their Universality and Education*. Berkeley: University of California Press.

Berndt D. J., and Clifford J. (1994). 'Using Dynamic Time Warping to Find Patterns in Time Series' in *Kdd Workshop*, Vol. 10, pp. 359–370.

Blasi D. E., Wichmann S., Hammarström H., Stadler P. F., and Christiansen M. H. (2016). 'Sound–Meaning Association Biases Evidenced Across Thousands of Languages', *Proceedings of the National Academy of Sciences*, 113: 10818–10823. https://doi.org/10.1073/pnas.1605782113

Campello R. J., Moulavi D., and Sander J. (2013). 'Density-Based Clustering Based on Hierarchical Density Estimates'. In: *Pacific-Asia conference on knowledge discovery and data mining*, pp. 160–172.

Carey S. (2009). *The Origin of Concepts*. New York: Oxford University Press.

Carr J. W., Smith K., Cornish H., and Kirby S. (2017). 'The Cultural Evolution of Structured Languages in an Open-Ended, Continuous World', *Cognitive Science*, 41: 892–923. https://doi.org/10.1111/cogs.12371

De Boer B., and Verhoef T. (2012). 'Language Dynamics in Structured form and Meaning Spaces', *Advances in Complex Systems*, 15: 1150021. https://doi.org/10.1142/S0219525911500214

Dingemanse M., Blasi D. E., Lupyan G., Christiansen M. H., and Monaghan P. (2015). 'Arbitrariness, Iconicity, and Systematicity in Language', *Trends in Cognitive Sciences*, 19: 603–615. https://doi.org/10.1016/j.tics.2015.07.013

Foutch B. K., Stringham J. M., and Lakshminarayanan V. (2011). 'A New Quantitative Technique for Grading Farnsworth D-15 Color Panel Tests', *Journal of Modern Optics*, 58: 1755–1763. https://doi.org/10.1080/09500340.2011.573881

Gibson E., Futrell R., Piantadosi S. P., Dautriche I., Mahowald K., Bergen L., and Levy R. (2019). 'How Efficiency Shapes Human Language', *Trends in Cognitive Sciences*, 23: 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Goldstein L. (2003). 'Emergence of Discrete Gestures' In: *Proceedings of the 15th international congress of phonetic sciences*, pp. 85–88.

Grice M., Ritter S., Niemann H., and Roettger T. B. (2017). 'Integrating the Discreteness and Continuity of Intonational Categories', *Journal of Phonetics*, 64: 90–107. https://doi.org/10.1016/j.wocn.2017.03.003

Gussenhoven C. (1999). 'Discreteness and Gradience in Intonational Contrasts', *Language and Speech*, 42: 283–305. https://doi.org/10.1177/00238309990420020701

Harnad S. (2003). 'Categorical Perception'.

Hawkins R. D., Franke M., Frank M. C., Goldberg A. E., Smith K., Griffiths T. L., and Goodman N. D. (2023). 'From Partners to Populations: A Hierarchical Bayesian Account of Coordination and Convention', *Psychological Review*, 130: 977–1016. https://doi.org/10.1037/rev0000348

Hawkins R. D., Goodman N. D., and Goldstone R. L. (2019). 'The Emergence of Social Norms and Conventions', *Trends in Cognitive Sciences*, 23: 158–169. https://doi.org/10.1016/j.tics.2018.11.003

Hockett C. (1960). 'The Origin of Speech', *Scientific American*, 203: 88–97.

Hofer M., Le T. A., Levy R., and Tenenbaum J. (2021). Learning evolved combinatorial symbols with a neuro-Symbolic generative model. arXiv:2104.08274 [cs], preprint: not peer reviewed. https://doi.org/10.48550/arXiv.2104.08274

Hofer M., and Levy R. P. Iconicity and structure in the emergence of combinatoriality Preprint. PsyArXiv. https://doi.org/10.31234/osf.io/vsjkt, 2019, preprint: not peer reviewed.

Kirby S., Cornish H., and 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*, 105: 10681–10686. https://doi.org/10.1073/pnas.0707835105

Kirby S., Tamariz M., Cornish H., and Smith K. (2015). 'Compression and Communication in the Cultural Evolution of Linguistic Structure', *Cognition*, 141: 87–102. https://doi.org/10.1016/j.cognition.2015.03.016

Kruskal J. B. (1964). 'Nonmetric Multidimensional Scaling: A Numerical Method', *Psychometrika*, 29: 115–129. https://doi.org/10.1007/BF02289694

Lawson R. G., and Jurs P. C. (1990). 'New Index for Clustering Tendency and its Application to Chemical Problems', *Journal of Chemical Information and Computer Sciences*, 30: 36–41. https://doi.org/10.1021/ci00065a010

Lewis D. K. (1969) *Convention: A Philosophical Study*. Hoboken, NJ: John Wiley & Sons.

Liberman A. M., Cooper F. S., Shankweiler D. P., and Studdert-Kennedy M. (1967). 'Perception of the Speech Code', *Psychological Review*, 74: 431–461. https://doi.org/10.1037/h0020279

Lieck R., and Rohrmeier M. (2021). 'Discretisation and Continuity: The Emergence of Symbols in Communication', *Cognition*, 215: 104787. https://doi.org/10.1016/j.cognition.2021.104787

Little H., Eryılmaz K., and De Boer B. (2017). 'Signal Dimensionality and the Emergence of Combinatorial Structure', *Cognition*, 168: 1–15. https://doi.org/10.1016/j.cognition.2017.06.011

Macuch Silva V., Holler J., Ozyurek A., and Roberts S. G. (2020). 'Multimodality and the Origin of a Novel Communication System in Face-to-Face Interaction', *Royal Society Open Science*, 7: 182056. https://doi.org/10.1098/rsos.182056

Mantel N. (1967). 'The Detection of Disease Clustering and a Generalized Regression Approach', *Cancer Research*, 27: 209–220.

Monaghan P., Mattock K., and Walker P. (2012). 'The Role of Sound Symbolism in Language Learning', *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 38: 1152–1164.

Monaghan P., Shillcock R. C., Christiansen M. H., and Kirby S. (2014). 'How Arbitrary is Language?', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369: 20130299. https://doi.org/10.1098/rstb.2013.0299

Morin O., Müller T. F., Morisseau T., and Winters J. (2022). 'Cultural Evolution of Precise and Agreed-Upon Semantic Conventions in a Multiplayer Gaming App', *Cognitive Science*, 46: e13113. https://doi.org/10.1111/cogs.v46.2

Motamedi Y., Wolters L., Naegeli D., Kirby S., and Schouwstra M. (2022). 'From Improvisation to Learning: How Naturalness and Systematicity Shape Language Evolution', *Cognition*, 228: 105206. https://doi.org/10.1016/j.cognition.2022.105206

Müller T. F., and Raviv L. (2021). 'Communication Games: Social Interaction in the Formation of Novel Communication Systems'.

Newport E. L. (1982). 'Task Specificity in Language Learning? Evidence from Speech Perception and American Sign Language', *Language Acquisition: The State of the Art*, 450–486.

Nölle J., Staib M., Fusaroli R., and Tylén K. (2018). 'The Emergence of Systematicity: How Environmental and Communicative Factors Shape a Novel Communication System', *Cognition*, 181: 93–104. https://doi.org/10.1016/j.cognition.2018.08.014

Pimentel T., McCarthy A. D., Blasi D. E., Roark B., and Cotterell R. (2019). Meaning to form: measuring systematicity as information. arXiv:1906.05906, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1906.05906

Raviv L., de Heer Kloots M., and Meyer A. (2021). 'What Makes a Language Easy to Learn? A Preregistered Study on How Systematic Structure and Community Size Affect Language Learnability', *Cognition*, 210: 104620. https://doi.org/10.1016/j.cognition.2021.104620

Raviv L., Meyer A., and Lev-Ari S. (2019). 'Compositional Structure can Emerge Without Generational Transmission', *Cognition*, 182: 151–164. https://doi.org/10.1016/j.cognition.2018.09.010

Rosch E. H. (1973). 'Natural Categories', *Cognitive Psychology*, 4: 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Sakoe H., and Chiba S. (1978). 'Dynamic Programming Algorithm Optimization for Spoken Word Recognition', *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26: 43–49. https://doi.org/10.1109/TASSP.1978.1163055

Sandler W., Aronoff M., Meir I., and Padden C. (2011). 'The Gradual Emergence of Phonological form in a New Language', *Natural Language & Linguistic Theory*, 29: 503–543. https://doi.org/10.1007/s11049-011-9128-2

Sandler W., Meir I., Padden C., and Aronoff M. (2005). 'The Emergence of Grammar: Systematic Structure in a New Language', *Proceedings of the National Academy of Sciences*, 102: 2661–2665. https://doi.org/10.1073/pnas.0405448102

Schanda J. (2007). *Colorimetry: Understanding the CIE System*. Hoboken, NJ: John Wiley & Sons.

Schouwstra M., and de Swart H. (2014). 'The Semantic Origins of Word Order', *Cognition*, 131: 431–436. https://doi.org/10.1016/j.cognition.2014.03.004

Shepard R. N. (1987). 'Toward a Universal Law of Generalization for Psychological Science', *Science*, 237: 1317–1323. https://doi.org/10.1126/science.3629243

Shepard R. N., and Cooper L. A. (1992). 'Representation of Colors in the Blind, Color-Blind, and Normally Sighted', *Psychological Science*, 3: 97–104. https://doi.org/10.1111/j.1467-9280.1992.tb00006.x

Sidhu D. M. (2025). 'Sound Symbolism in the Lexicon: A Review of Iconic-Systematicity', *Language and Linguistics Compass*, 19: e70006. https://doi.org/10.1111/lnc3.v19.1

Székely G. J., Rizzo M. L., and Bakirov N. K. (2007). 'Measuring and Testing Dependence by Correlation of Distances'.

Theisen C. A., Oberlander J., and Kirby S. (2010). 'Systematicity and Arbitrariness in Novel Communication Systems', *Interaction Studies*, 11: 14–32. https://doi.org/10.1075/is

Verhoef T. (2012). 'The Origins of Duality of Patterning in Artificial Whistled Languages', *Language and Cognition*, 4: 357–380. https://doi.org/10.1515/langcog-2012-0019

Verhoef T., Kirby S., and De Boer B. (2016). 'Iconicity and the Emergence of Combinatorial Structure in Language', *Cognitive Science*, 40: 1969–1994. https://doi.org/10.1111/cogs.2016.40.issue-8

Verhoef T., Roberts S. G., and Dingemanse M. (2015). 'Emergence of Systematic Iconicity: Transmission, Interaction and Analogy'. in Proceedings of the *37th Annual Meeting of the Cognitive Science Society*.

Verhoef T., Walker E., and Marghetis T. (2016). 'Cognitive Biases and Social Coordination in the Emergence of Temporal Language'. in *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.

Woods K. J., Siegel M. H., Traer J., and McDermott J. H. (2017). 'Headphone Screening to Facilitate Web-Based Auditory Experiments', *Attention, Perception, & Psychophysics*, 79: 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Xu J., Dowman M., and Griffiths T. L. (2013). 'Cultural Transmission Results in Convergence Towards Colour Term Universals', *Proceedings of the Royal Society B: Biological Sciences*, 280: 20123073. https://doi.org/10.1098/rspb.2012.3073

Zaslavsky N., Kemp C., Regier T., and Tishby N. (2018). 'Efficient Compression in Color Naming and Its Evolution', *Proceedings of the National Academy of Sciences*, 115: 7937–7942. https://doi.org/10.1073/pnas.1800521115

Zuidema W., and De Boer B. (2018). 'The Evolution of Combinatorial Structure in Language', *Current Opinion in Behavioral Sciences*, 21: 138–144. https://doi.org/10.1016/j.cobeha.2018.04.011

Zuidema W., and Westermann G. (2003). 'Evolution of An Optimal Lexicon Under Constraints from Embodiment', *Artificial Life*, 9: 387–402. https://doi.org/10.1162/106454603322694834