Proceedings of the Annual Meeting of the Cognitive Science Society

# Emergence of vowel-like organization in a color-based communicatino system

Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Roberts, Gareth
Clark, Robin

2018

# Emergence of vowel-like organization in a color-based communication system

**Gareth Roberts (gareth.roberts@ling.upenn.edu)**
Department of Linguistics, 3401-C Walnut Street
Philadelphia, PA 19104 USA

**Robin Clark (rclark@sas.upenn.edu)**
Department of Linguistics, 3401-C Walnut Street
Philadelphia, PA 19104 USA

## Abstract

Vowel systems exhibit organization, and several theoretical accounts have been proposed to explain this. A prominent account explains organization in terms of maximizing the dispersion of vowels, increasing acoustic perceptibility while reducing articulatory effort. This implies modality-independence, but leaves open questions about the extent to which dispersion is driven by articulatory or acoustic pressures. We investigated whether vowel-like organization would emerge in a novel visual communication system in the laboratory, in which participants took turns to send color signals to communicate a set of animal referents by moving their fingers around a color space. We manipulated the extent to which sender and receiver needs were aligned. Overall, systems exhibited significant levels of dispersion; participants also took into account receiver needs, withconsequences for the structure of the resulting systems.

**Keywords:** language; phonology; experimental; communication game; experimental semiotics

The phonologies of natural languages exhibit a high degree of organization in their choice and deployment of phonemes. This can be clearly demonstrated with vowels. A vowel is produced by allowing air from the lungs to pass through the vocal tract, with only low levels of constriction. Vowel quality is varied by varying the shape of the vocal tract, principally by moving the tongue and lips. This leads to variation in formant frequencies, prominent concentrations of acoustic energy. Several formants matter for speech perception, but the first and second are the most important. Traditionally, the vowel phonemes of a language are plotted in a two-dimensional space with the x axis corresponding to the second formant (F2, shown as increasing in frequency from right to left), and the y axis to the first formant (F1, shown as increasing from top to bottom); see Figure 1. These values correspond well enough to tongue position that vowels located towards the top left of the space (i.e., with high F2 values and low F1 values) are standardly referred to as high front vowels; that is, these are vowels produced by raising the tongue towards the top front of the mouth (for a more detailed introduction to the phonetics of vowels see Ladefoged & Johnson, 2015). The purpose of this paper is to present a novel experimental approach to understanding the origins of, and constraints on, the organization of such spaces.

That vowel systems exhibit substantial organization has long been recognized (Liljencrants & Lindblom, 1972; Schwartz, Boë, Vallée, & Abry, 1997; de Boer, 2000). While the vowelspace is continuous, the vowel phonemes used in a given language are not simply distributed at random across it. Rather, they tend to be dispersed relatively efficiently. In a three-vowel system, for instance, it is extremely likely that one vowel will be a high front vowel /i/, one will be a high back vowel /u/, and the third will be a low vowel /ɑ/ or /a/. Larger systems tend to exhibit similarly efficient structure. Several different categories of theory have been proposed to account for this observation (for reviews see de Boer, 2001; Vaux & Samuels, 2015). While some accounts focus on features of individual vowels (e.g., classical "markedness"-based accounts; Jakobson & Halle, 1956; Chomsky & Halle, 1968), others focus on the relationship between vowels in a system. In *dispersion theory*, a particularly prominent account, optimization of the system is taken to be driven by the functional pressures of minimizing effort (in particular that of the speaker) while maximizing the perceptual contrast between vowels (Liljencrants & Lindblom, 1972; Lindblom, 1986; de Boer, 2001).

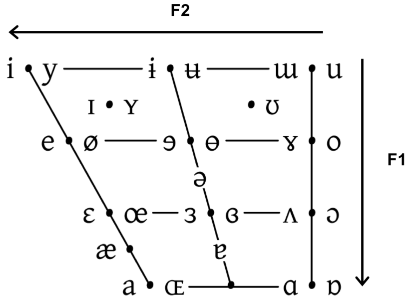

Figure 1: Chart of the vowels of the world's languages (based on charts produced by the International Phonetic Association), with F1 and F2 axes indicated. The chart corresponds roughly to the mouth of a speaker facing left. For each pair of vowels, the vowel on the right is produced with lip rounding.

In fact, most accounts of vowel space organization – with markedness-based accounts a notable exception – claim essentially that vowel space self-organization is driven by two basic demands: increasing acoustic perceptibility and reducing articulatory load (Stevens & Keyser, 2010; Mrayati, Carré, & Guérin, 1988). These demands compete in some cases. It is notable, for instance, that if dimensions beyond the first and second formant are taken into account, vowel systems do not in fact make maximal use of the resources available to them. By nasalizing one of the vowels in a three-vowel system and pharyngealizing another, for instance, their mutual distinctiveness could be increased (cf. de Boer, 2001, p. 15). This is uncommon, however, and the explanation for that would appear to be that incorporating these extra dimensions increases articulatory complexity for the speaker, and does not become necessary from the point of view of perceptual distinctiveness until the number of vowels in a system gets sufficiently large (Lindblom & Maddieson, 1988).

Nevertheless, the details of how these demands relate to each other are still not fully clear, and in many respects they are hard to disentangle. It was noted above that the frequencies of the first and second formants map somewhat well to tongue position, meaning that – in this respect at least – there is a reasonably consistent relationship between articulatory space and acoustic space. This makes it hard to distinguish speaker-driven constraints from listener-driven ones. It might seem intuitive that dispersion is likely to be driven by perceptibility rather than articulatory ease, but this is in fact not obvious. The edges of the space are advantageous from an articulatory point of view, as they are easier to find than arbitrary points within the space (for this reason, the corners of the space are particularly advantageous).[1] To put the question another way: If acoustic distinctiveness were reduced rather than increased by moving to the edges of the articulatory space, would we see the same pattern? If perceptibility is the main driving force, we should expect not to; if dispersion is driven more by articulatory constraints, then we should expect such a system to look similarly dispersed to real-world systems. Answering this question by manipulating the acoustic and articulatory space is in principle possible, but poses severe difficulties, not least reducing the influence of participants' own natural language phonologies. In this paper, therefore, we present an experiment in which participants communicated visually, using fingers as articulators to produce colors as analogs of vowels.

Our study has more than one goal. The first goal is to establish a new experimental approach to investigating vowel space organization. As this implies, it is anticipated that the general approach presented here will be applied to many relevant questions. The second goal concerns the specific question to which we applied the approach. This question has two parts. First, we tested whether a novel visual communication system would, through cooperative communication, begin to exhibit organization at greater than chance level. Second, we manipulated the extent to which the interests of the "speaker" aligned with those of the "listener" and tested whether, when those interests were not aligned, systems would take that into account and be organized differently.

Our paradigm draws on a set of approaches that were primarily developed, and have become an increasingly standard method, for investigating the emergence and cultural evolution of language. These approaches, which Galantucci (2009) dubbed *Experimental Semiotics*, involve human participants learning or creating novel communication systems in the laboratory (for a review, see Galantucci, Garrod, & Roberts, 2012), and they involve a social dimension lacking in traditional artificial-language-learning experiments. In experimental-semiotic studies, participants either engage in a communication task (e.g., Galantucci, 2005; Sneller & Roberts, 2018) or learn a miniature language based on the output of other participants (e.g., Kirby, Cornish, & Smith, 2008; Verhoef, Kirby, & de Boer, 2014). Taking an experimental-semiotic approach has a number of advantages. Analogues to change that would take many years in natural language can be observed in miniature languages over very short time periods; factors that cannot be manipulated outside the laboratory, or cannot be manipulated in natural languages even within the laboratory, can be manipulated in miniature languages with relative ease. By investigating processes of change in non-linguistic communication systems, particularly using novel signaling media, researchers can reduce the influence of the participants' own languages (for further discussion of the advantages of this approach, see Galantucci & Roberts, 2012). Although few experimental-semiotic studies (to our knowledge) have investigated the *organization* of combinatorial units in phonological spaces (de Boer & Verhoef, 2012, is something of an exception), several have investigated the *emergence* of combinatorial units from continuous signals (e.g., Roberts & Galantucci, 2012; Roberts, Lewandowski, & Galantucci, 2015; Verhoef et al., 2014; Del Giudice, 2012). Our study was influenced by these: Participants played a cooperative signaling game, sending color signals to each other to communicate a set of animal referents. We manipulated the way different colors could be produced and measured the organization of the resulting systems.

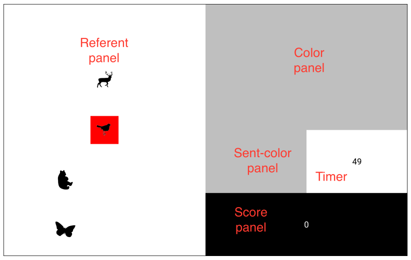

Figure 2: Sender's screen. Labels are for clarity and were not shown to participants.

# Method

## Participants

Sixty University of Pennsylvania students (34 female), none of whom suffered from color-blindness, participated in pairs for course credit.[2]

---

[1] A comparison can be made with consonants; stop consonants involve bringing articulators into complete contact, while fricatives involve holding them close enough that the air passes through with turbulence. Stop consonants are acquired earlier by children than fricatives and are more common in the world's languages (Ladefoged & Maddieson, 1996).

[2] Owing to a software error, other demographic data such as age and handedness were not recorded. However, their distribution is

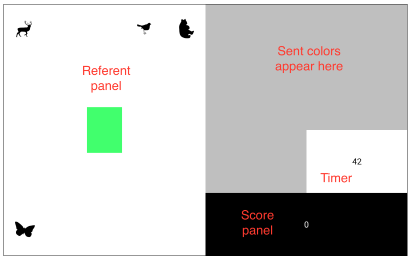

Figure 3: Receiver's screen. Labels are for clarity and were not shown to participants.

## Materials

Participants sat in separate cubicles, each with a computer (a mid-2014 Apple iMac), running custom-designed software (written in Python and Kivy), and a wireless multitouch trackpad (a 2009 Apple Magic Trackpad, measuring 13.01cm by 13.13cm). Participants could not see each other from their cubicles or hear each other easily.

## Procedure

Pairs of participants played a cooperative communication game, taking turns to be *Sender* and *Receiver*.[3] Each participant (henceforth *player*) in a pair sat in a separate cubicle and saw a screen divided vertically into two halves. (For the most part, the screen looked much the same whether the player was Sender or Receiver; Figures 2 and 3). In the left half of the screen, the *referent panel*, a set of *referents* were displayed (black animal silhouettes, a subset of those used by Roberts & Galantucci, 2012, Figure 4).[4] The top right half of the screen, the *color panel*, appeared gray as default, but would change color depending on the behavior of the Sender. The same was true of a smaller section immediately below it, the *sent-color panel*, which was also gray by default and took up a quarter of the width of the screen as a whole and a quarter of the height. (See *Signaling medium* Section below for a description of how the color panel and the sent-color panel worked.) To the right of the sent-color panel, a timer was displayed on a white background. Below this, taking up half the width of the screen, was a *score panel* displaying the pair's joint score against a black background.

The referent panel differed slightly for the Sender and the Receiver. First, the referents were not in the same places (and were redistributed randomly each round). Second, no referent was ever in the center of the Receiver's referent panel; the Sender, on the other hand, always had one referent in the center, against a red background (Figure 2). Third, the Receiver had a green cursor that could be moved around the referent panel by using the arrow keys on the computer keyboard (Figure 3). The Sender had no such movable cursor. The Sender's task was to convey to the Receiver which referent they had in the center of their panel by sending colors to the Receiver (see *Signaling medium* Section below), and the Receiver's task was to move the cursor to the correct referent and press enter. Both players would then receive feedback: The background of the correct referent would turn red for the Receiver and the background of the chosen referent would turn green for the Sender. This happened whether or not the Receiver chose correctly. If the Receiver did choose correctly, the pair would score one point; their total point score was displayed in the score panel at the bottom of the screen. After players started to do well at signaling the referents, more were added, in groups of four, up to a total of twelve. This would occur if, for all referents in the referent panel, the Receiver had selected them correctly at least 75% of the time over the previous four rounds in which they had occurred (cf. Roberts et al., 2015).

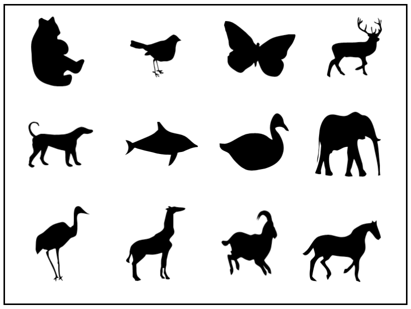

Figure 4: Referents used in the experiment. The top row appeared at the beginning of the game; as players became more successful, the other rows were added in turn.

A round lasted 20s in total, with feedback lasting an additional 2s. If the Receiver had not chosen a referent by the time the 20s were up, the pair scored no point for that round. Whatever the outcome of the round, the players would swap roles for the following round. The game lasted for 80min in total, and would finish at the end of the current round when the 80min mark had been passed. At the start of the experiment, players played four practice rounds that differed from the ordinary rounds in three ways: First, they lasted 60s; second, the players' score from these rounds did not carry over into the normal rounds; third, players were reminded at the start of each round whether they were Sender or Receiver. Beyond being told to move a finger around the pad and observe the screen, and to hold their finger down for 1s to send a color, players were not instructed how to use the signaling medium, but rather had to explore it on their own.

---

not expected to have deviated substantially from that of the wider undergraduate population at Penn.

[3] It was important for our question that participants both have an opportunity to be Sender and Receiver. Had this not been the case, any differences between conditions could be explained in terms of a failure on the part of the Sender to appreciate the Receiver's needs. This approach also had the advantage of greater ecological validity.

[4] Roberts and Galantucci (2012) used 20 referents in total; we used a 12-referent subset of theirs in order to give participants time to refine their signaling systems. Given the time available, continuing to add referents until there were twenty would have meant that systems would be in a constant state of flux.

## Signaling medium

To convey to the Receiver which referent to select, the Sender could send colors. The Sender could do this by moving one finger around on the trackpad. This would produce a color in the color panel on the top right of their screen (but not the Receiver's screen), which would change in real time depending on the coordinates of the Sender's finger. If the Sender took their finger off the pad or touched the pad with more than one finger, the color panel would appear gray. If the Sender held their finger in place on the trackpad for 1s or longer, the same color would appear for 2s both on the Receiver's color panel and on the Sender's sent-color panel. This was the only means by which the Sender could send information to the Receiver.[5] A Sender could send as many colors as they liked, within the time available (20s).

The relationship between the Sender's finger position and the color produced was based on an RGB color space, with each color composed of red, green, and blue components, the contribution of each ranging from 0 to 1 (e.g., the vector [1,0,0], where the digits indicate the red, green, and blue components respectively, would correspond to a bright red color). The basic value for one of the three components increased from 0 to 1 as the Sender's finger moved from right to left on the pad, while another decreased from 1 to 0 in the same direction; the third color component increased as the finger moved vertically. Which color corresponded to which direction was counterbalanced, but for any trial, the exact center of the pad corresponded to the vector [0.5, 0.5, 0.5]. If vertical position corresponded to the blue component, then placing the finger in the middle of the top edge of the pad would produce a mixture of red and green [0.5, 0.5, 0], while the middle of the bottom edge would produce a mixture of red, green, and blue, with blue predominating: [0.5, 0.5, 1]. Players were not in fact exposed precisely to the basic color values described here; instead, the values were modified in a way that varied between two conditions. The details of this are described in the *Conditions* section.

## Conditions

There were two conditions. In the *Bright-edge* condition the basic color values described above were altered depending on how close the Sender's finger was to the center of the pad (Figure 5). This was done by multiplying the color component values by a modifier that ranged from 0 to 1. The modifier was calculated as $d/d_{oe}$, where $d$ equals the Euclidean distance between the Sender's finger and the center of the space and $d_{oe}$ equals the distance from the center of the space to the outer edge. This meant that colors towards the edges of the space were brighter, and therefore likely to be clearer for the Receiver to distinguish. Since the edges of the pad were also easier to find reliably for the Sender, the pressures acting on the Sender and Receiver were therefore relatively aligned in this condition.

In the *Bright-band* condition this was not the case. Here, an imaginary line was drawn 30% of the way in from the edge of the pad. Between the real edge of the pad and this "inner edge", the modifier was calculated as $1-(d/d_{oe})$. Once the Sender's finger crossed the inner edge, however, the modifier changed to $d/d_{ie}$, where $d_{ie}$ is the distance from the center of the pad to the inner edge. This meant that the colors got brighter as the Sender's finger moved away from the center of the pad, but then began abruptly to get darker again. The most convenient parts of the pad for the Sender to select reliably were still along the outer edge of the pad, but the easiest colors to distinguish for the Receiver were closer to the inner edge. The inner edge was in no way marked on the pad or screen; it became apparent to the Sender as they moved their finger around the pad and observed the effect.

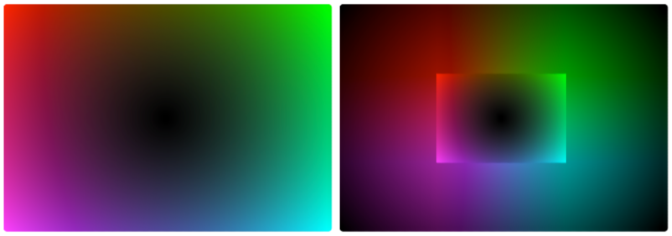

Figure 5: Example color spaces for Bright edge and Bright band conditions respectively. Note that participants never saw the space itself, only individual colors.

## Dependent variables

For each player, we looked at the last successful signal for every referent for which the player could be determined to have established a signal. In terms of the game, this meant that the referent had been successfully communicated in at least three of the last four times it had occurred. This gave us a set of "words" that that player had in their system, each of which was composed of a set of units – colors that had been chosen to be sent. Each unit consisted of an x and a y coordinate corresponding to the player's finger position when the color was sent.[6] We pooled all units across a player's words and produced a "phoneme inventory". As in natural language, it was assumed that players might reuse phonemes between words. We therefore trimmed each inventory by comparing phonemes; if the distance between any two of them was less than 5% of the maximum distance available, one was removed at random. This left an inventory ranging in size from

---

[5] In this respect, our study differs from earlier experimental-semiotic work on combinatorial systems, which were primarily concerned with investigating the emergence of atomic units from continuous media. Participants in those studies were thus not provided with pre-ordained means of producing units. This means that identifying how such units might be constituted is itself a challenging task (Roberts & Galantucci, 2012). Because of this, and because we were concerned not with the emergence of such units, but how they become organized, our task forced subjects to select units from a continuous space, thereby simplifying our analysis while still maintaining a continuous signal space from which units could be drawn.

[6] For the purposes of the analysis presented here, we will focus only on the finger position, ignoring the color coordinates produced.

7 to 19 phonemes (Mean = 11.97). Figure 6a shows an example 12-phoneme inventory from the Bright-edge condition.

We then measured the following variables. In each case, the mean value for a pair forms the basis of the results reported below.

**Phoneme inventory size.** The number of phonemes in an inventory.

**Dispersion.** Two measures of dispersion were used:

**Mean pairwise distance.** The mean Euclidean distance between all pairs of phonemes. This gives a measure of how well spread out the phonemes are within the space. This was then divided by the maximum possible distance to give a number between 0 and 1.

**Mean distance from center.** The mean Euclidean distance between the center of the space and the phonemes. This was then divided by the distance from the center to the corner to give a number between 0 and 1.

**Number of referents.** The number of referents for which a "word" had successfully been established.

**Success index.** This measure was based on how many words a player successfully established and how fast they did so. For every round of a given game, we counted how many referents each player had an established word for (see above) at that point. We then calculated a success index as $(\sum_1^{n_r} s)/12n_r$, where $n_r$ is the number of rounds and the numerator is a cumulative count of $s$, the number of successfully established words in a given round.[7]

We present comparisons between conditions for all variables below. For the second and third variables (mean pairwise distance and mean distance from center) we also generated random artificial data to provide a baseline. We did this by first counting the number of units in each inventory, before trimming (see above) had occurred. Then we generated a random inventory of phonemes of the same size, and trimmed that. Then we measured the mean pairwise distance and mean distance from center for the artificial data. This was repeated 10,000 times, and we counted the number of times the values for the variables in the randomly generated data were equal to or higher than the real data. If this occurred often, it would suggest that the real data did not exhibit particular high levels of dispersion. In fact, the values in the baseline trials did not in any one of the 10,000 replications equal or exceed the values for the real data in either condition.

## Results

Communication systems in the Bright-edge condition employed slightly more phonemes (M = 12.77; SD = 2.15) than in the Bright-band condition (M = 11.17; SD = 1.81; $t(27.19) = 2.2$, $p = 0.04$). There was a positive correlation between the number of phonemes and the number of referents successfully communicated, $r = 5.5$, $n = 30$, $p = 0.0018$.

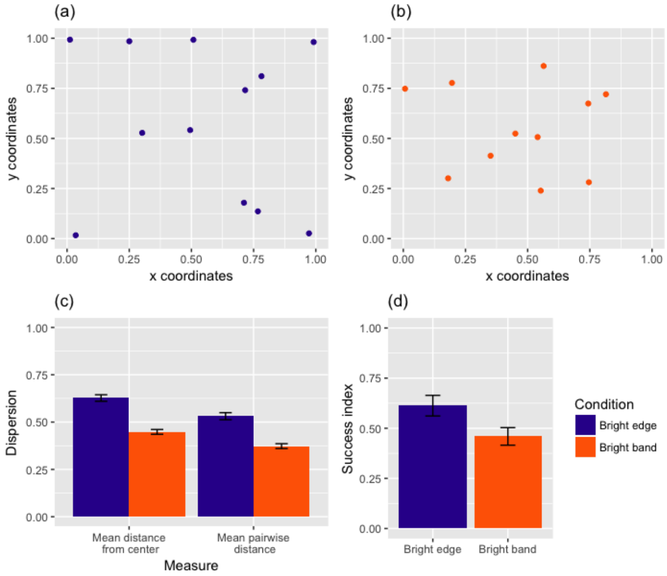

Figure 6: (a) Example inventory from Bright-edge condition. Coordinates are normalized by dividing by the width and height of the space; (b) Example (normalized) inventory from Bright-band condition. (c) Mean dispersion; (d) Mean success. Error bars show 95% CI.

Players in the Bright-band condition appear to have taken into account the Receiver's interests: Phonemes in this condition were less far from the center of the space (M = 0.53; SD = 0.05) than in the Bright-edge condition (M = 0.63; SD = 0.05; $t(27.82) = 5.25$, $p < 0.0001$; Figure 6c). The pairwise distance was also lower (M = 0.37; SD = 0.036) than in the Bright-edge condition (M = 0.45; SD = 0.035; $t(27.99) = 5.79$, $p < 0.00001$; Figure 6d).

Players in the Bright-band condition also found the game harder (Figure 6b). They established successful signals for a mean of 8.3 referents (SD = 2.04), and their mean success index was 0.46 (SD = 0.12). The figures for the Bright-edge condition were significantly higher in each case: 10.83 referents (SD = 1.72; $t(27.2) = 3.68$, $p = 0.001$), with a mean success index of 0.61 (SD = 0.14; $t(27.303) = 3.15$, $p = 0.004$).

Success also correlated with both Mean distance from center ($r(28) = 0.39$, $p = 0.032$) and Mean pairwise distance ($r(28) = 0.39$, $p = 0.032$)[8]

## Discussion

We set out to answer two questions: First, whether a collaboratively constructed visual communication system would

---

[7] It should be noted that no player could actually score 1, as that would require them to have successfully communicated all twelve referents several times before the start of the game. While this could be accounted for, this would complicate the calculation, which we did not deem necessary for a relative measure of success.

[8] An earlier version of this paper reported no correlation; this was an error based on correlating dispersal with number of referents – a more indirect proxy for success – rather than the success index.

exhibit organization analogous to that of a vowel space; second, whether receivers' interests with respect to perceptibility would be allowed to trump senders' interests with respect to ease of production. The answer to both questions was a clear yes. Our data provide support for dispersion-based accounts of phonological organization (as opposed to accounts based on language-specific notions of markedness), suggesting that the organizational principles involved are modality-independent. There are some limitations of the current study. In natural language the articulatory space maps to the acoustic space in a particular way that has likely been subjected to evolutionary forces. The mapping of the trackpad space to the color space in our experiment is rather more arbitrary and has clearly not been subjected to the same kind of evolutionary forces. This has advantages (as our paradigm provides greater space for manipulating variables, and helps shield our data from natural-language influence), but in future work it would be worthwhile taking into account more specific details of natural-language speech production. Our study also did not take into account several factors that have been suggested as playing a role in the organization of vowel spaces, not least generational transmission (Vaux & Samuels, 2015), which might be expected to amplify the biases operating in our experiment (Kalish, Griffiths, & Lewandowsky, 2007). There are well established methods that can be used to investigate this in future work (Kirby et al., 2008; Verhoef et al., 2014). However, we consider that, in this paper, we have presented a novel and useful experimental paradigm for investigating the cultural evolution of phonological spaces, which can be easily adapted to answer a number of questions.

## Acknowledgments

We thank members of the Cultural Evolution of Language Lab for their work gathering data for this project.

## References

Chomsky, N., & Halle, M. (1968). *The sound patterns of English*. Cambridge, MA: MIT Press.

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

de Boer, B. (2001). *The origins of vowel systems*. Oxford: Oxford University Press.

de Boer, B., & Verhoef, T. (2012). Language dynamics in structured form and meaning spaces. *Advances in Complex Systems*, *15*(3–4), 1150021 (20 pages).

Del Giudice, A. (2012). The emergence of duality of patterning through iterated learning: Precursors to phonology in a visual lexicon. *Language and Cognition*, *4*(4), 381–418.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, *29*(5), 737–67.

Galantucci, B. (2009). Experimental Semiotics: A new approach for studying communication as a form of joint action. *Topics in Cognitive Science*, *1*(2), 393–410.

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

Galantucci, B., & Roberts, G. (2012). Experimental Semiotics: An engine of discovery for understanding human communication. *Advances in Complex Systems*, *15*(3–4), 1150026.

Jakobson, R., & Halle, M. (1956). *Fundamentals of language*. The Hague: Mouton.

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, *14*(2), 288–294.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Ladefoged, P., & Johnson, K. (2015). *A course in phonetics* (Seventh ed.). Stamford, CT: Cengage Learning.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*(4), 839–62.

Lindblom, B. (1986). Phonetic universals in vowel systems. *Experimental phonology*, 13–44.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. *Language, Speech, and Mind*, 62–78.

Mrayati, M., Carré, R., & Guérin, B. (1988). Distinctive regions and modes: A new theory of speech production. *Speech Communication*, *7*(3), 257–286.

Roberts, G., & Galantucci, B. (2012). The emergence of duality of patterning: Insights from the laboratory. *Language and Cognition*, *4*(4), 297–318.

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, *141*, 52–66.

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). Major trends in vowel system inventories. *Journal of Phonetics*, *25*, 233–253.

Sneller, B., & Roberts, G. (2018). Why some behaviors spread while others don't: A laboratory simulation of dialect contact. *Cognition*, *170C*, 298–311.

Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*(1), 10–19.

Vaux, B., & Samuels, B. (2015). Explaining vowel systems: Dispersion theory vs. natural selection. *The Linguistic Review*, *32*(3), 573–599.

Verhoef, T., Kirby, S., & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, *43*, 57–68.