UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Neural measures of sensitivity to a culturally evolved space-time language: shared biases and conventionalization

Permalink
https://escholarship.org/uc/item/8ck4z1dk

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

Authors
Verhoef, Tessa
Walker, Esther
Marghetis, Tyler
et al.

Publication Date
2018

eScholarship.org Powered by the California Digital Library
University of California

Neural measures of sensitivity to a culturally evolved space-time language: shared biases and conventionalization

Tessa Verhoef (t.verhoef@liacs.leidenuniv.nl)
Leiden Institute of Advanced Computer Science, Niels Bohrweg 1
2333 CA Leiden, Netherlands

Esther Walker (e1walker@ucsd.edu)
Department of Cognitive Science, 9500 Gilman Drive
La Jolla, CA 92093 USA

Tyler Marghetis (tmarghet@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E 10th St
Bloomington, IN 47405 USA

Seana Coulson (scoulson@ucsd.edu)
Department of Cognitive Science, 9500 Gilman Drive
La Jolla, CA 92093 USA

Abstract

When asked to convey temporal concepts such as 'yesterday' and 'tomorrow' via movements of a dot on a vertical bar, American undergraduates utilize analogical mappings between spatial and temporal concepts. Previous work has revealed two different strategies, hypothesized to require differing amounts of artificial language exposure to learn. Different pairs of participants, when interacting about these time concepts, all settled on the same association between spatial magnitude and temporal duration, with larger movements used to convey temporal intervals of greater duration. However, the association between particular spatial locations and temporal concepts such as 'past' and 'future', elicited much more arbitrary solutions, where the mappings differed across pairs of participants. These findings suggested that the duration mapping might be driven by mostly shared, initial cognitive biases, while contrasting mappings for past/future result more clearly from extensive linguistic interaction. Here we tested whether the brain responds differently to duration mappings as compared to direction mappings by recording participants' EEG as they learn a mini-language that includes both kinds. ERPs time locked to English words elicited larger amplitude N400 and P600 when they did not match the preceding signal than when they did match. The P600 results were larger and more robust for the duration than the direction stimuli, suggesting participants were more sensitive to violations of the duration mapping scheme. These data support our hypothesis that people have a cognitive bias for the duration mappings that supports their early emergence in the development of a semiotic system.

Keywords: analogy; artificial language learning; iterated learning; language comprehension; language evolution; metaphor; N400; P600

Introduction

How does the neural basis of language learning relate to the structure of language itself? This is a question that requires spanning the brain, communicative behaviors, and the dynamics of cultural systems. Studies of language evolution, including computational simulations (Griffiths & Kalish, 2007; Hare & Elman, 1995; Kirby & Hurford, 2002) and behavioral experiments (Kirby, Cornish & Smith, 2008; Smith & Wonnacott, 2010; Verhoef, 2012), suggest an account of linguistic structure as both 'shaped by the brain' (Deacon, 1997; Christiansen & Chater, 2008) and refined by cultural evolution (Kirby & Hurford, 2002; Kirby, Cornish & Smith, 2008; Kirby, Griffiths, & Smith, 2014). On this view, the structured nature of language reflects its status as a complex adaptive system (Steels, 1997; Beckner, Blythe, Bybee, Christiansen, Croft, Ellis, Holland, Ke, Larsen-Freeman & Schoenemann, 2009), constrained by individual neural biases and the demands of cultural transmission.

Although the conclusion of such experimental work is often that linguistic systems adapt to cognitive constraints during social transmission, previous work has, to the best of our knowledge, never directly measured the brain response to the languages that evolve in laboratory experiments (but see Lumaca & Baggio (2016) for a similar approach in the domain of auditory processing and the emergence of tonal/musical systems), relying instead on indirect behavioral measures of processing demands or complexity (e.g., reproduction accuracy, entropy as in Kirby, Cornish & Smith, 2008; Verhoef, 2012). In the current study we will directly assess the brain response to an emerging semiotic system, offering insight into the mechanisms by which such systems become more structured and learnable. Whereas previous studies have used ERPs to study the acquisition of experimenter-generated artificial grammars (e.g. Silva, Folia, Hagoort & Petersson, 2017), to our knowledge the current study is the first to examine the brain response to language-like systems that have emerged from actual communicative interactions and transmission between real people in the lab.

Our focus is the domain of *time* and its iconic communication via the use of the spatial domain. The use of language about space to convey concepts of time is a well-established linguistic phenomenon ("long duration", "back in the past", Lakoff & Johnson, 1980). However, the mechanisms by which these space-time mappings emerge is a matter of continuing debate (Winter, Marghetis & Matlock, 2015). Some argue for the importance of innate neural representations (Walsh, 2003), others for analogical reasoning based on shared structure (Gentner, 1983), and still others for conceptual metaphors learned from experience (Lakoff & Johnson, 1980). While each of these mechanisms may play a role, none on their own can account for the combination of universality and variability that we find in the languages of the world, a pattern that requires explanatory mechanisms operating at multiple timescales (Núñez & Cooperrider, 2013). Consequently, time is a particularly good test case for understanding interactions between cultural evolution and initial individual biases.

Here we address the interplay between culture and cognition by exploring whether factors that characterize the emergence of novel meanings in social settings also predict their learnability by subsequent individual learners. In doing so, we build on previous work on the emergence of space-time mappings through communication games (Verhoef, Walker & Marghetis, 2016). In this study, pairs of participants (in separate rooms) communicate about temporal concepts using only a novel, spatial signaling device: a vertical bar on a touch screen (Fig. 1a), that records and replays brief movement patterns of a bubble. The results indicated that for some parts of the emerging structure, there was a clear involvement of mostly **shared, initial biases**, while other parts seemed to result more clearly from a process of negotiation and **social coordination**. For instance, spatial length was used consistently to indicate temporal duration, as all dyads used *larger* regions of vertical space to indicate *longer* durations (Fig. 1b), likely the result of shared cognitive biases. Further, while the bar's vertical location (top/bottom) or the direction of movement (up/down) was typically used to indicate deictic time (past/future), the particular mappings varied from dyad to dyad (Fig. 1c, d), indicating more idiosyncratic solutions resulting from conventionalization through coordination and interaction.

In a follow-up study, Verhoef, Walker & Marghetis (2018) let these initial systems evolve further through *iterated* communication games (Tamariz, Cornish, Roberts & Kirby, 2012; Verhoef, Roberts & Dingemanse, 2015), where the signals that were developed by one pair were used as initial training for the next, before they started interacting. Transmission chains of 8 interacting pairs were created in which the signaling systems evolved over time. A fully systematized, regular language emerged in this experiment as the result of repeated transmission. In the present study, we expose participants to such a language that emerged in the Verhoef et al. (2018) study, while we record their brainwaves with EEG.

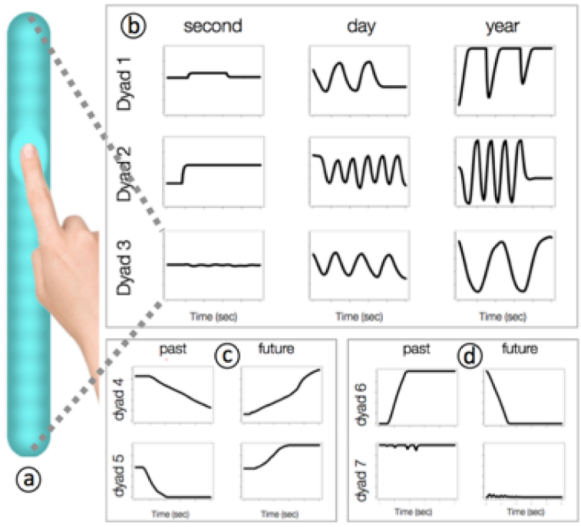


Figure 1: (a) Signaling device, (b) Consistent use of spatial length to communicate relative duration, (c) and (d) Contrasting mappings for past/future.

Expectations

The findings of the earlier behavioral study by Verhoef et al. (2016) suggested that the spatial length to duration mapping might be driven by strong neural biases, while contrasting direction mappings for past/future emerge from a social negotiation process, and are more conventionalized. In the present study, we record EEG as participants view signals from the language followed by English words that either match (Correct) or mismatch (Violations) their meanings. We predict participants' brain response will reflect greater sensitivity to the meanings that involve duration mappings than for the more arbitrary direction ones. Two ERP components are of particular interest, the N400 component, whose amplitude is highly sensitive to the semantic fit between a word and its context, and the P600, whose amplitude is known to be enhanced by a variety of linguistic violations, ranging from rule-based grammatical violations to semantic anomalies. If participants are sensitive to the meanings of the signals in the miniature language, both the N400 and the P600 would be expected to be larger following violations than correct pairings. Greater neural sensitivity to one sort of violation – for example, if participants display a cognitive bias for learning duration mappings – might be registered by larger amplitude violation effects on the N400 and/or the P600.

Methods

Participants

Participants were 15 undergraduate students at the University of California, San Diego (average age = 19.7 years; 4 men), who participated as part of a course requirement for a class in Cognitive Science, Psychology, or

Linguistics. All were right-handed fluent English speakers with normal or corrected-to-normal vision, and no reported psychiatric or neurological disorders.

Materials

Signals and time words were a subset of those used in Verhoef et al. (2016), see Figure 2. These pairings could either be "correct" or made up of one of four types of violations: day-year (duration) switch, vertical (direction) switch, random far or random close. In a day-year switch, the English translation for a temporal duration signal is swapped with a different duration signal (e.g., the signal for "year" is presented with the word "day" or the signal for "tomorrow" is presented with the words "next year"). In a direction switch, the English translation of "past" or "future" is paired with a signal that goes in the opposite direction of the intended mapping (e.g., a downward signal for "past" is presented with the word "future" or an upward signal for "after" is presented with the word "before"). For the random violations, signals were paired with random English translations that were either semantically close to the intended meaning (e.g., year → this year; future → after) or semantically far from the intended meaning (e.g., day after → last year; yesterday → year after), based on a semantic distance measure that was previously defined by Verhoef et al. (2016).

Procedure

After being prepped with EEG electrodes (see next section), participants entered a dimly lit, sound-attenuated chamber. They were told that for each trial, they would view a short signal, followed by a potential English translation of the signal. Participants were asked to respond by pushing a button to indicate whether or not the translation was correct ("yes" or "no") and would receive feedback.

Each trial (see figure 3) began with a fixation cross in the center of the screen (random duration between 500 and 1000 ms). Then, a moving signal was presented for 1.5 seconds on a vertical blue bar against a white screen, followed by the English translation, which remained on the screen for 500 ms. Then, a question mark appeared on the screen, signaling the participants to respond. Participants' responses triggered a positive or negative feedback tone that enabled learning.

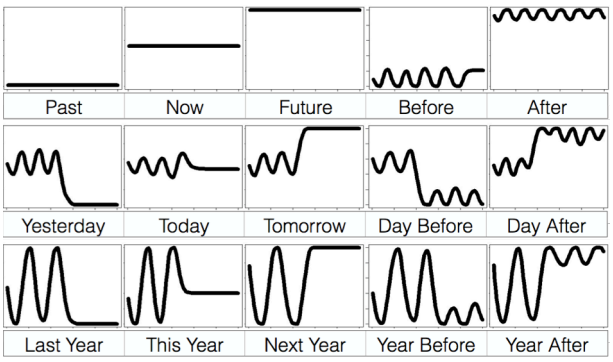


Figure 2: The signal-word mappings used. The y-axis represents location of the bubble on the vertical bar, while the x-axis represents time.

Figure 3: The procedure for one trial: first a fixation cross is shown, then the signal appears showing the movement, this is followed by the English translation, then a question mark is shown asking for a participant response and finally a feedback tone is played.

Each participant completed at least two rounds of 96 trials, with a maximum of 3 rounds. Each of the 16 items was presented 6 times per round: 3 times in a correct signal-word pairing, and 3 times in one of the four violation types.

EEG Recording and Analysis

EEG was recorded using 29 tin electrodes, organized in the International 10-20 configuration in a mesh cap, and referenced to the left mastoid during recording. Two electrodes were placed at the outer canthi to record eye movements, and one was placed below the right eye to detect blinks. EEG was recorded using an SA Instruments bioelectric amplifier with a high pass filter of 0.01Hz and a low pass filter of 100Hz, and was digitized online at 512Hz.

ERPs were time locked to the English words following each signal. They were examined for artifacts due to movement and channel blockage. An average of 16.92% (sd=7.9) of trials were rejected due to artifacts. Mean amplitude measurements 250-450ms (N400) and 500-700ms (P600) were taken from electrodes F3, F4, FC3, FC4, C3, C4, CP3, CP4, P3, P4, O1, and O2 and subjected to repeated measures ANOVA as outlined below.

Results

Initial omnibus analyses of each interval involved repeated measures ANOVA with factors Round(First, Second), Stimulus(Duration, Direction), Status(Correct, Violation), and Electrode(12 levels). A snapshot of results can be found in Figure 4, which shows the overall status effect (middle panel), collapsed across round and stimulus type, as well as comparisons of ERPs elicited by violations and correct pairings for duration (top) and direction (bottom) stimuli. Relative to correct pairings, violations tended to elicit enhanced N400 and more positive P600. The distribution of the status effect (violation minus correct pairing) across the scalp is illustrated in the accompanying colored heads in Figure 4.

N400

Our initial analysis revealed a reliable main effect of Status, $F(1,14) = 7.670$, $p < 0.05$, ges = 1.559e-02, as violations elicited more negative ERPs than words in correct pairings. We also observed a trend for interactions between Round and Electrode, $F(10,140) = 2.657$, and between Stimulus, Round, and Electrode, $F(10, 140) = 1.901$, neither of which survived epsilon adjustment (Huynh & Feldt, 1976). Because stimulus type was of a priori interest, we conducted separate follow up analyses of the Direction and Duration conditions with factors Round (First, Second), Status (Correct, Violation), and Electrode (12 levels). Analysis of the Direction condition revealed only a trend for an interaction between Round and Electrode, $F(10, 140) = 2.991$, that did not survive epsilon adjustment. Analysis of the Duration condition revealed a reliable main effect of Status, $F(1,14) = 6.066$, $p < 0.05$, ges = 1.911e-02. Duration violations elicited ERPs that were on average 1.3 microvolts more negative than did the correct pairings.

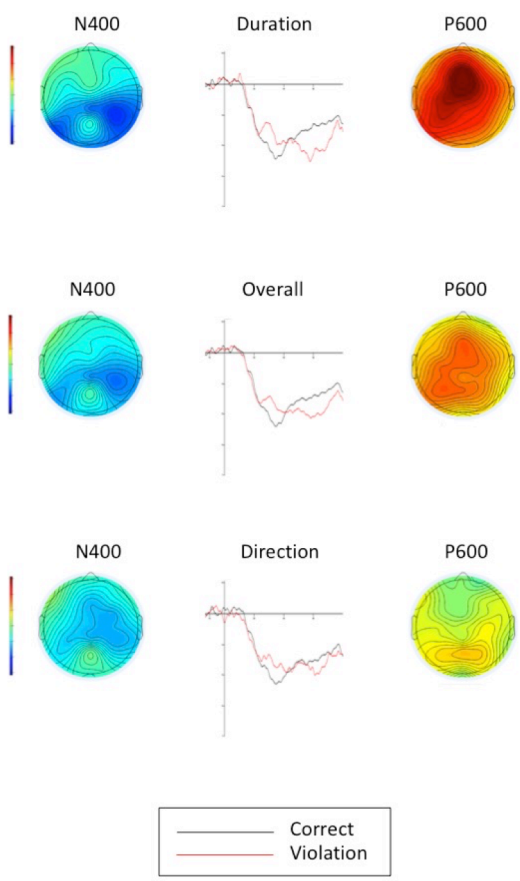


Figure 4: ERPs from a composite electrode elicited by English words in correct pairings in black versus violations in red. Following the convention in the ERP language literature, negative voltage has been plotted upwards.

P600

Our initial analysis revealed a reliable main effect of Status, $F(1,14) = 6.591$, $p < 0.05$, ges = 0.004, qualified by an interaction with Stimulus, $F(1,14) = 13.693$, $p < 0.05$, ges = 0.007, and an interaction between Status and Round, $F(1,14) = 4.921$, $p < 0.05$, ges = 0.009. To explore the source of these interactions, we conducted separate follow-up analyses of data elicited by Direction stimuli and for Duration stimuli. These repeated measures ANOVAs included factors Round (First, Second), Status (Correct, Violation), and Electrode (12 levels). Follow-up analysis of the direction condition revealed no reliable effects. By contrast, follow up analysis of the Duration condition revealed a main effect of Status, $F(1,14) = 11.778$, $p < 0.05$, ges = 5.077e-02, as the violations elicited larger P600 than did the correct translations.

Task Performance

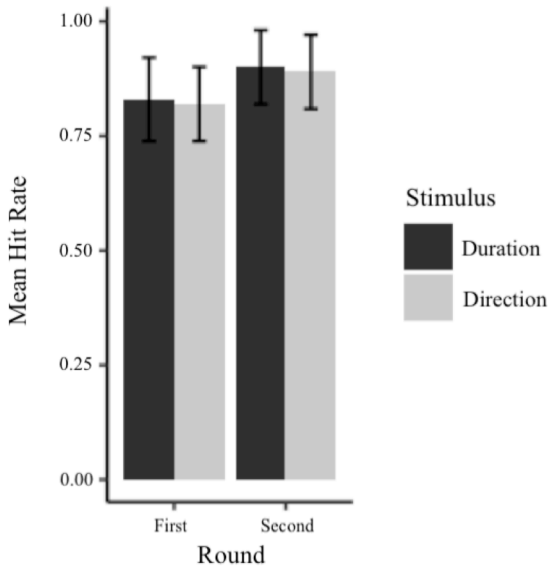
Performance on the task was assessed by compiling each participant's percentage of hits (when they responded 'match' for a correct mapping) and false alarms (when they responded 'mismatch' for a correct mapping) for the Duration and Direction stimuli in each round. Hit rates were subjected to repeated measures ANOVA with factors Round and Stimulus. Performance improved in the second round, $F(1, 14) = 9.63$, $p < 0.05$, ges = 0.005, but Stimulus was not significant either as a main effect or in interaction with Round. Similarly, analysis of false alarm rates revealed a main effect of Round, $F(1, 14) = 20.88$, $p < 0.05$, ges = 0.15, as participants made fewer false alarms in the second round than they did in the third round. Stimulus did not reliably affect false alarm rates, though examination of the data suggests this may be due to limitations in the power of the study to detect subtle behavioral effects (ges was 0.05 for Stimulus and 0.01 for the Stimulus x Round interaction) for the Stimulus effect was 0.05, with a relatively small sample size (n=15).

Discussion

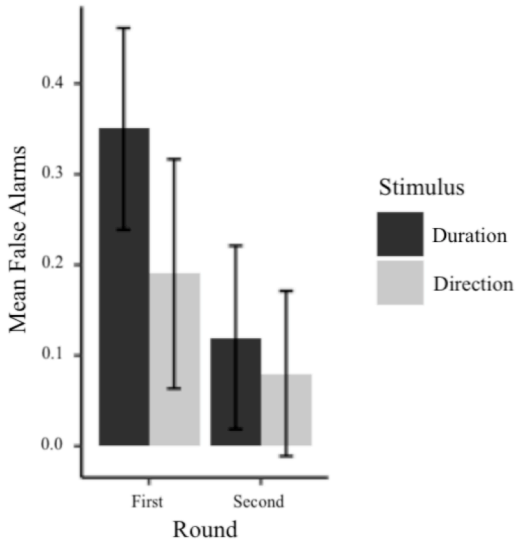
In the present study, we investigated the influence of cognitive biases on emerging space-time mappings in an artificial language by explicitly measuring neural responses to learning trials. Previous work, using an experimental paradigm to investigate the cultural evolution and conventionalization of space-time mappings in artificial languages, suggested (but could not prove) the presence of two complimentary drives towards structure. The emergence of mappings between temporal duration and spatial extent seemed to be driven by strong initial biases, shared by most participants. By contrast, mappings between vertical spatial locations and direction to past/future events seemed to be more arbitrary and idiosyncratic.

Here we tested whether the brain responds differently to duration mappings as compared to direction mappings by recording participants' EEG as they learned an artificial language that includes both of these kinds of mappings.

ERPs time locked to English words that did not match the preceding signal elicited larger amplitude N400 and P600 than when they did. This suggests that participants noticed the systematicity in the language they were exposed to and were sensitive to the meanings of its signals. Although behavioral measures of performance suggested participants



(a)



(b)

Figure 5: Average Proportion of Hits, where participants chose 'match' for a correct mapping (a) and False Alarms, where participants chose 'mismatch' for correct mappings (b) in each of the conditions. Error bars depict 95% confidence intervals.

did equally well on direction and duration mappings, violations of duration mappings were larger and more robust than for direction ones. These data are in keeping with our hypothesis that the duration mappings that emerged in our dyadic iterated learning experiment, result from cognitive biases shared by many individuals, whereas the more variable direction mappings are more idiosyncratic, and become conventionalized in a more arbitrary fashion. Because participants likely had more variable expectations regarding the nature of the direction mappings, their brain response to such violations was similarly more variable. The lack of a consistent brain response across the group of participants would explain why the direction violations elicited low amplitude effects that were not statistically significant.

Our results are similar to a finding by Lockwood, Hagoort, & Dingemanse (2016), in which Dutch participants were asked to learn sound-symbolic words from a language they don't speak: Japanese. These words were iconic in the sense that there was an intrinsic connection of resemblance between the form and meaning. Participants were exposed to either correct pairings or meanings paired with an opposite word, violating the iconic link. The results showed that people had a harder time learning the opposite pattern, and this was also reflected in their ERPs at test. In our artificial space-time language, the duration mappings can be seen as having iconicity, where the analogical mapping between spatial magnitude and temporal duration creates a perceptual resemblance between form and meaning. Even though we did not try to explicitly teach our participants the opposite mapping, as in Lockwood et al. (2016), we did find a clear neural sensitivity for these more iconic mappings as opposed to the more arbitrary direction mappings to past and future.

More broadly, our results show that, when learning novel mappings between the domains of space and time, the brain distinguishes between those regularities that seem to derive from shared initial biases and those that derive from convention. In the context of language emergence and the evolution of structure, these results enable a novel integration of neural measures and experimental methods that model processes of cultural evolution. Through this combination, the interactions between neural biases and sociocultural experience in the course of learning and transmitting a novel mini-language can be investigated in more detail. Here, we presented neural evidence for the proposal that linguistic structure is both shaped by the brain (Deacon, 1997; Christiansen & Chater, 2008) and refined by cultural evolution (Kirby & Hurford, 2002; Kirby, Cornish & Smith, 2008; Kirby, Griffiths & Smith, 2014) and that both are needed to explain the combination of universality and variety in languages around the world (Núñez & Cooperrider, 2013). This study opens up exciting avenues of future work in which the learnability of languages that originate in iterated learning paradigms can be empirically tested with the use of EEG measurements. This will allow us to be more precise about how neural biases influence

learning and transmission in the emergence of structure in language and what the role is of different interacting mechanisms. This will help to more definitively establish whether and how cultural evolution enables language to be more compatible with the neural architecture.

Acknowledgments

We are grateful to Tania Delgado, Megan Bardolph, Emily Sereno and Ankhit Pandurangi for help with data collection. This work is supported by the Netherlands Organisation for Scientific Research (NWO Veni grant 275-89-031 to T.V.)

References

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59(s1), 1–26.

Christiansen, M. H. & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(5):489–509.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. WW Norton & Co Inc.

Gentner, D. (1983). Structure mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.

Griffiths, T. & Kalish, M. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive Science*, 31(3), 441–480.

Hare, M. & Elman, J. L. (1995). Learning and morphological change. *Cognition*, 56(1), 61–98.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of educational statistics*, *1*(1), 69-82.

Kirby, S. & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In *Simulating the evolution of language*, pp. 121–147. Springer.

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.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current opinion in neurobiology*, *28*, 108-114.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lockwood, G., Hagoort, P., & Dingemanse, M. (2016). How iconicity helps people learn new words: Neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra: Psychology*, *2*(1).

Lumaca, M., & Baggio, G. (2016). Brain potentials predict learning, transmission and modification of an artificial symbolic system. *Social cognitive and affective neuroscience*, *11*(12), 1970-1979.

Núñez, R. E., & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Topics in Cognitive Science*, 17, 220–229.

Silva, S., Folia, V., Hagoort, P., & Petersson, K. M. (2017). The P600 in implicit artificial grammar learning. *Cognitive science*, *41*(1), 137-157.

Smith, K. and Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, 116, 444–449.

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of communication*, 1(1), 1–34.

Tamariz, M., Cornish, H., Roberts, S. & Kirby, S. (2012) The effect of generation turnover and interlocutor negotiation on linguistic structure. In Scott-Phillips, T., Tamariz, M., Cartmill, E. A., & Hurford, J. (Eds.) *The evolution of language: Proceedings of the 9th international conference (evolang9)*. World Scientific, (pp. 555-556).

Verhoef, T. (2012) The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, 4(4), 357-380.

Verhoef, T., Roberts, S. G., Dingemanse, M. (2015) Emergence of systematic iconicity: Transmission, interaction and analogy. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D. & Maglio, P. P. (Eds.) *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2481-2486) Austin, TX: Cognitive Science Society.

Verhoef, T., Walker, E. & Marghetis, T. (2016). Cognitive biases and social coordination in the emergence of temporal language. In Papafragou, A., Grodner, D., Mirman, D., & Trueswell, J.C. (Eds.). *Proceedings of the 38th Annual Conference of the Cognitive Science Society.* (pp. 2615-2620) Austin,TX: Cognitive Science Society.

Verhoef, T., Walker, E. & Marghetis, T. (2018). Cognitive biases and cultural evolution in the emergence of space-time mappings in language. In Cuskley, C., Flaherty, M., Little, H., McCrohon, L., Ravignani, A. & Verhoef, T. (Eds.): The Evolution of Language: Proceedings of the 12th International Conference (EVOLANGXII).

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Topics in Cognitive Science*, 7, 483–488.

Winter, B, Marghetis, T., and Matlock, T. (2015). Of metaphors and magnitudes: Explaining cognitive interactions between space, time, and number. *Cortex,* 64, 209-224.