Brief article

Perceptual categories enable pattern generalization in songbirds

Jordan A. Comins [a], Timothy Q. Gentner [a,b,c,*]

[a] Department of Psychology, University of California, San Diego, United States
[b] Neurosciences Program, University of California, San Diego, United States
[c] Kavli Institute for Brain and Mind, La Jolla, United States

ARTICLE INFO

ABSTRACT

Since Chomsky's pioneering work on syntactic structures, comparative psychologists interested in the study of language evolution have targeted pattern complexity, using formal mathematical grammars, as the key to organizing language-relevant cognitive processes across species. This focus on formal syntactic complexity, however, often disregards the close interaction in real-world signals between the structure of a pattern and its constituent elements. Whether such features of natural auditory signals shape pattern generalization is unknown. In the present paper, we train birds to recognize differently patterned strings of natural signals (song motifs). Instead of focusing on the complexity of the overtly reinforced patterns, we ask how the perceptual groupings of pattern elements influence the generalization pattern knowledge. We find that learning and perception of training patterns is agnostic to the perceptual features of underlying elements. Surprisingly, however, these same features constrain the generalization of pattern knowledge, and thus its broader use. Our results demonstrate that the restricted focus of comparative language research on formal models of syntactic complexity is, at best, insufficient to understand pattern use.

1. Introduction

Theories of syntax are central to understanding language. As a result, many comparative psychologists interested in the study of language evolution have targeted pattern complexity, using formal mathematical grammars, as the key to organizing language-relevant cognitive processes across species (Berwick, Beckers, Okanoya, & Bolhuis, 2012; Berwick, Okanoya, Beckers, & Bolhuis, 2011). In identifying similarities and differences between human and nonhuman animals in relation to pattern recognition and production, some researchers have concluded that comparative models are inadequate because "human language sentences are potentially unbounded in length and structure, limited only by extraneous factors, such as short-term memory or lung capacity" (Berwick et al., 2011). This focus on formal syntactic complexity, however, disregards the close interaction in real-world signals between the structure of a pattern and its constituent elements as well as core biological and cognitive constraints intrinsic to temporal processing and, therefore, language. Others have argued that comparative studies are essential to the study of language precisely because they showcase how biological and cognitive mechanisms interact with dynamic real-world signals to tune pattern perception mechanisms crucial to aspects of language (Kiggins, Comins, & Gentner, 2012; Margoliash & Nusbaum, 2009). The latter perspective proposes to study language and its evolution in the context of the principles of organismal biology (Margoliash & Nusbaum, 2009), whereas the former posits these questions in the domain of mathematical formalisms specifically unburdened by such restrictions (Berwick et al., 2011, 2012).

* Corresponding author. Address: UCSD, Department of Psychology, 9500 Gilman Drive MC0109, La Jolla, CA 92092-0109, United States. Tel.: +1 858 822 6763.
E-mail address: tgentner@ucsd.edu (T.Q. Gentner).

What role does syntactic pattern complexity play in language origin? Is the capacity for humans to perceive *mildly context-sensitive language* constructions, such as "John Mary Peter Jane lets help teach swim" (see Berwick et al., 2011), or even *context-free language* constructions, like "the starling the cats want was tired" (again, Berwick et al., 2011) what marks the divide between a nervous system that can support language from one that cannot? Or are these grades of temporal pattern complexity *extraneous* to the minimally sufficient requirements for language? Regardless of formal computational models that might be instantiated to perceive a given pattern, the broader utility of any pattern is constrained by its generalizability. That is, can knowledge of a pattern gained in one context be employed to recognize the same pattern in other contexts.

To explore this idea, we examined how perceptual constraints shape pattern generalization in European starlings, *Sturnus vulgaris*, a species of songbird. Starlings are expert auditory sequence learners both in the laboratory and during natural song development. Acoustically, starling song appears as a patterned sequence of distinct motifs, where each 200–1000 ms long motif comprises multiple shorter notes presented in a stereotyped pattern. Although motif repertoires between starlings are largely unique, all motifs can be grouped into four general classes by their broad acoustic characteristics (Eens, Pinxten, & Verheyen, 1988). Perceptually, these classes form open-ended natural categories for starlings (Braaten, 2000), and individual motifs within categories can be readily differentiated (Meliza, 2011). Starlings attend to the temporal patterning of notes within motifs (Gentner, 2008) and to the temporal patterning of motifs in songs (Gentner & Hulse, 1998). Further, starlings accurately classify and generalize artificial motif patterns that take the forms AABB and ABAB (Gentner, Fenn, Margoliash, & Nusbaum, 2006), where sets A and B represent two of four natural motif categories (rattles and warbles) in starling song.

2. Methods

2.1. Subjects

Eight (*N* = 8) European starlings (*S. vulgaris*) wild-caught in southern California served as subjects. We controlled for neither age nor sex, though all subjects had full adult plumage when acquired and, thus, were at least one year old. From the time of capture until use in this study, subjects were housed in a large mixed sex, conspecific aviary with *ad libitum* access to food and water. The photoperiod in the aviary and the testing chambers followed the seasonal variation in local sunrise and sunset times.

2.2. Stimuli

Patterned stimuli were made using motifs from the song recordings of one adult male starling captured near Baltimore, Maryland (for details, see Gentner & Hulse, 1998). Motifs can be broadly classified into four classes based on their spectrotemporal features: whistles, warbles, rattles and high frequency events (Adret-Hausberger & Jenkins,

1988). We extracted eight unique (non-repetitive) warble and rattle motifs from the songs of one male. Both warbles and rattles possess a variable acoustic structure, with rattles further characterized by a broadband click train. The rattle and warble motif categories follow ethologically-relevant and psychologically-salient boundaries (Braaten, 2000; Eens et al., 1988). Using these two classes of motifs, we built two patterns of motif sequences of the form XXYY or XYXY. While the pattern constituents and groupings (i.e. motifs and motifs classes) are ethologically-relevant, the overall pattern forms XXYY/XYXY are not.

To construct XXYY and XYXY patterned stimuli, we divided the motifs into two sets, labeled A and B. For half of the subjects (*n* = 4), motif membership in set A and B preserved natural acoustic (perceptual) category boundaries (i.e., 8 warbles for set A; 8 rattles for set B; see Fig. 1). For the other half of subjects (*n* = 4) motif membership in sets A and B conflicted with the natural perceptual boundaries (i.e., 4 warbles, 4 rattles forset A; 4 warbles, 4 rattles forset B; see Fig. 2C and D).

Subjects were trained using 32 (out of a possible 16,384) XXYY and XYXY patterned stimuli (8 of the explicit form AABB, 8 BBAA, 8 ABAB and 8 BABA). Within each subset of 8 explicit training patterns (e.g. AABB), each motif appeared exactly once in each possible position, and no motif appeared twice in the same sequence. No differences were observed between the mean duration of warbles and rattles (independent samples *t*-test: $t = 0.98$; $df = 14$; $p = 0.34$), or in the durations of XXYY and XYXY patterns in the group where the natural perceptual boundaries A/B motif were preserved (independent samples *t*-test: $t = 0.0$; $df = 30$; $p = 1.00$), or pseudo-randomized (independent samples *t*-test: $t = 0.0$; $df = 30$; $p = 1.00$). Thus, stimulus length cannot be used to recognize patterns.

2.3. Apparatus

Fig. 2A illustrates the operant apparatus where starlings learned to classify the training stimuli. Each subject was held in a small weld-wire cage containing an operant panel. On the panel, a centrally located response port was a PVC-housed opening. Inside of this opening was an IR receiver and transmitter used to detect when the bird broke the plane of the response port with its beak. This 'pokehole' design allows starlings to probe the apparatus with their beaks, in a manner akin to their natural appetitive foraging behavior. A remotely controlled hopper, positioned behind the panel, moved the food within and beyond the subject's reach beneath the opening. Acoustic stimuli were delivered through a small full-range audio speaker mounted behind the panel and out of the subject's view. The sound pressure level inside all chambers was calibrated to the same standard broadband noise signal. Full details for all the mechanical components of the apparatus, audio interface, digital I/O control hardware, and custom software are available upon request.

2.4. Shaping

Upon initially entering the operant chamber, we provided each subject with unrestricted access to the food
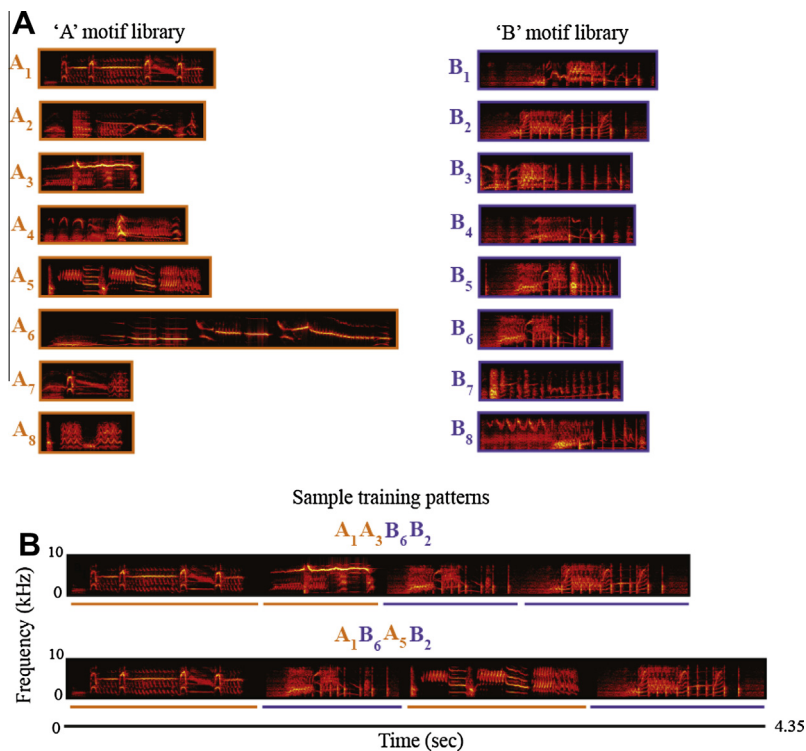
Fig. 1. (A) The 8 warbles and 8 rattles used to construct XXYY and XYXY pattern stimuli. While both warbles (left – orange) and rattles (right – dark blue) are heterogeneous in their overall spectrotemporal features, rattles additionally possess broad-band click trains not found in warbles. (B) Two out a total of 32 training patterns built from these motif libraries are shown below. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

hopper. Following acclimation to eating from this device, the hopper was lowered beyond the subject's reach. Next subjects were placed on an autoshaping routine (Brown & Jenkins, 1968). Several times per minute, we presented the subject with a blinking LED in the center response port followed by 2-s access to the food hopper. This process recurred until the subject acquired a key-peck response. From then, the lowered food hopper would only be engaged if the subject pecked the blinking LED in the center port. Subjects repeated this behavior for 100 trials. After the completion of this phase, the center LED ceased blinking, requiring subjects to peck at the darkened center port to raise the food hopper. Following 100 such trials, pecking the center port initiated the playback of an acoustic stimulus where subjects earned food rewards in accordance with standard go/no-go training procedures described below.

2.5. Classification training

Fig. 2B demonstrates the operant procedure. We trained subjects to classify sets of starling song stimuli abiding by two different patterning rules. To initiate a trial, subjects pecked a darkened center port which triggered the playback of a training stimulus. Once stimulus presentation concluded, the animal could either peck the darkened center port once more (termed a 'go' response) or not (termed a 'no-go' response). A peck in response to one set of stimuli

(S+ training set) was positively reinforced with 2-s access to the food hopper. A peck in response to the other set of training stimuli (S-training set) was punished by extinguishing the operant box light at least 10 s, thereby prohibiting food hopper access. An absence of a response to either the S+ or S− stimuli yielded no operant consequence. Correction trials were available in the experiment during only training phases. For these trials, a key-peck response within the response window following an S− stimulus not only extinguished the house lights as usual, but the program would also continue to present the same S− stimulus until the subject abstained from giving a key-peck response. Correction trials were discontinued for this experiment once the animal achieved our criterion for successful classification performance during the initial training phase (criterion: 3 successive $d' \geqslant 1.0$). The stimulus exemplar presented on any given non-correction trial was sampled randomly with replacement from the pool of all stimuli the animal was learning to classify. To evaluate classification accuracy, we labeled a response to an S+ stimulus and the withholding of a response to an S− stimulus as correct. Conversely, withholding a response to an S+ stimulus and responding to an S- stimulus were labeled incorrect. Subjects could freely peck at the center response port throughout stimulus presentation, but only the first response within a 2-s response window beginning at stimulus offset triggered reinforcement or punishment. Subjects were on a closed economy during training, with daily ses-
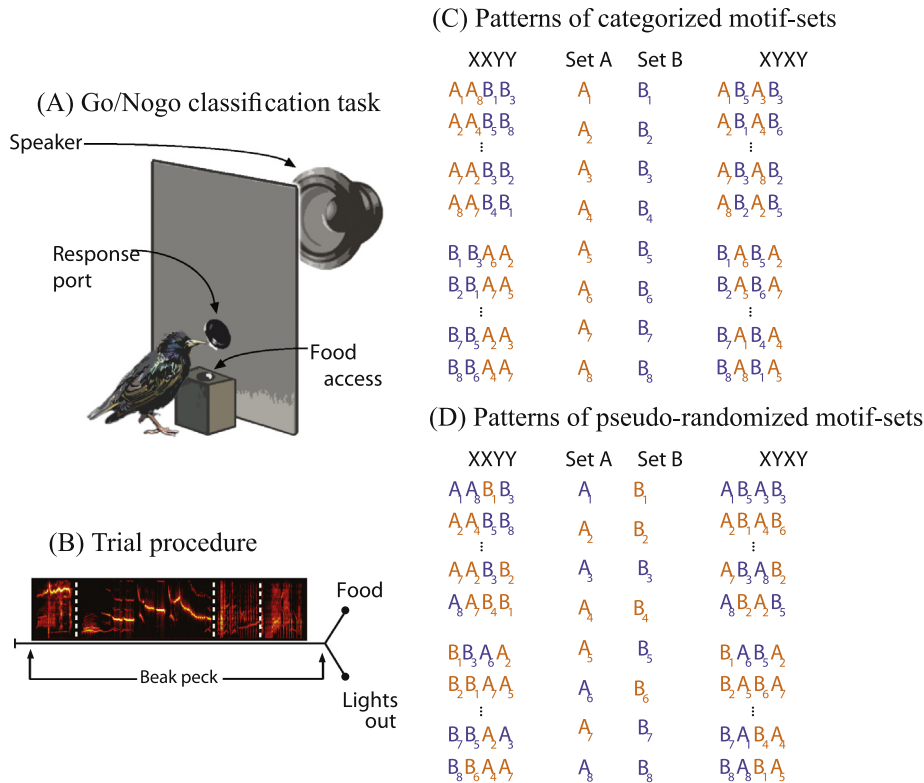
Fig. 2. (A) Schematic of operant training apparatus and (B) procedure. Visualization of pattern stimuli built from 2 sets of 8 motifs that either preserve natural acoustic categorical boundaries (C) or do not (D). Orange letters designate warbles, while dark blue designates rattles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sions lasting from sunrise to sunset, and each subject could run as few or as many trials as they were able. Food intake was monitored daily to ensure each subject's well-being. The inter-trial interval was a minimum of 2-s. Water was always available. All procedures were approved by the UCSD institutional animal care and use committee.

2.6. Procedure

To quantify learning during training we used a performance criterion of 3 consecutive 100-trial blocks where d' scores exceeded 1.0. Subjects remained on training stimuli after reaching this learning threshold to ensure stable behavioral performance prior to generalization testing. We assessed the subjects' ability to generalize knowledge of the XXYY and XYXY patterns presented during initial training using an abrupt and complete transfer to 500 novel motif sequences that followed the training patterns. Each of the 500 test sequences was built using the training motifs in sets A and B assigned to that subject, and was presented only once. The entire generalization test was exactly 500 trials. During transfer, we maintained all reinforcement contingencies as in training, except that no correction trials were used.

We used d-prime (d') to measure bias-free sensitivity of baseline and test pattern classification as follows: $d' = z(H - \varepsilon) - z(F + \varepsilon)$, where $z(H - \varepsilon)$ and $z(F + \varepsilon)$ are the z-scores of the proportions of responses to S+ and S− stim-

uli adjusted by a factor of $\varepsilon = 0.001$ to set response rate bounds $(0, 1)$, respectively. To compare generalization performance between the two groups, we normalized d' by dividing every subject's d' on each of their five 100-trial transfer blocks by their mean performance across the five 100-trial blocks preceding transfer.

3. Results

Initial pattern learning was similar for both groups. We observed no significant differences between groups in acquisition rate (number of 100-trial blocks until $d' > 1.0$ for 3-consecutive blocks; $t = -2.14$, df = 6, $p = 0.076$; Fig. 3A). More importantly, all subjects learned to classify the training patterns with accuracy significantly above chance ($p < 0.02$ all cases, t-test per bird for last 5 100-trial blocks of training, where chance $d' = 0$; Fig. 3B).

Surprisingly, pattern generalization differed qualitatively between groups. Subjects that learned patterns of categorically-organized motifs maintained accurate classification across the 500 novel patterns ($p < 0.05$ all cases, t-test per bird relative to chance; Fig. 3C). In contrast, none of the subjects that learned patterns with pseudo-randomly grouped motifs performed above chance during the transfer (t-test for each bird; in three cases: $p > 0.3$; one subject below chance; $p = 0.044$; Fig. 3C). Consistent with these differences, normalized generalization perfor-
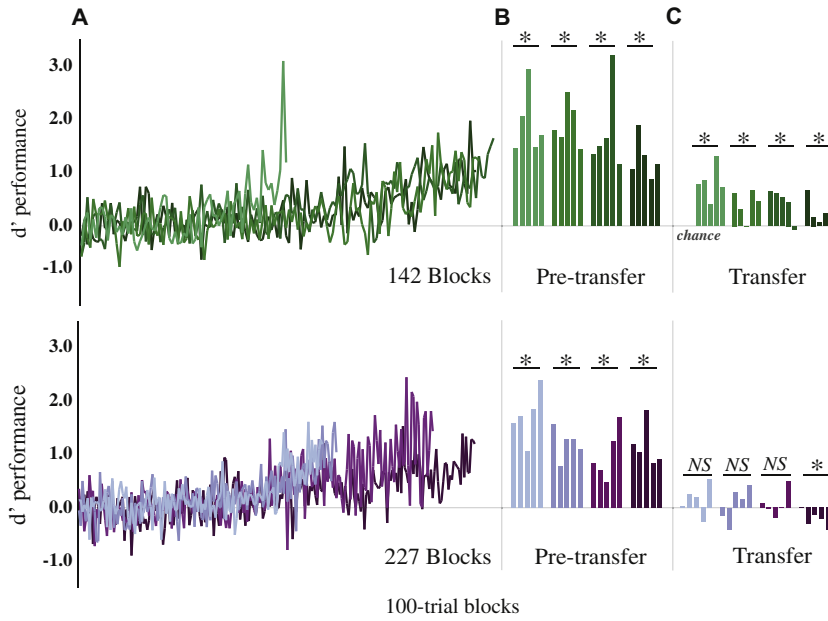
Fig. 3. (A) Training performance (d') on XXYY/XYXY patterns, where d'=0 represents chance (each subject is a different saturation). (B) Each bar shows performance of one subject across the final five 100-trial blocks of training; green (top) shows subjects trained with categorically organized motifs, purple (bottom) shows data for subjects trained with pseudorandomly organized motifs. Here, every subject was well-above chance performance, indicating reliable learning of XXYY and XYXY patterns. (C) Generalization performance across the five 100-trial blocks of novel XXYY and XYXY patterns differed dramatically between groups, with every subject trained on categorically organized motif patterns displaying above chance generalization and none of the subjects classifying patterns of pseudo-randomly organized motifs doing so. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

mance (see methods) for subjects trained with the categorically organized motifs was significantly better than that for subjects trained with the pseudo-randomly organized motifs ($F(1,6) = 13.67$, $p = 0.01$; nested rmANOVA; Fig. 4).

4. Discussion

The results of this simple experiment are clear. Reshuffling pattern elements across the boundaries of natural acoustic categories leaves pattern learning intact, but fully blocks pattern generalization. While pattern learning and pe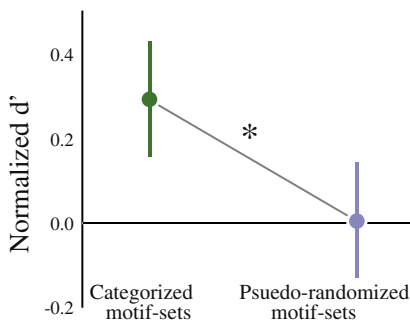rception in a restricted context (i.e., explicit training sequences) appears agnostic to the perceptual grouping of underlying elements, such grouping constrains generalization of pattern knowledge. Thus, regularity in the perceptual structure of the pattern elements is a salient feature of pattern perception in non-humans and is essential to the broader use pattern knowledge.

Although this experiment was not designed to test the perceptual strategies that each group of subjects used to acquire pattern knowledge during baseline training, the results are instructive nonetheless. Given that subjects trained with the pseudo-randomized motif groups failed to generalize to novel pattern exemplars, the most parsimonious strategy for their recognition of the baseline sequences is rote memorization. In contrast, rote memorization cannot account for generalization to the novel patterns by the group working with the categorically well-defined motifs. Although we observed no statistically significant difference in the acquisition rates between subjects in the two groups, there was a trend for those learning patterns with the pseudo-category motifs to take longer, and it is likely that with larger groups sizes these differences would reach statistical significance, consistent with a different learning strategy. Understanding the learning strategy that enables successful generalization will require further experiments, but it clearly involves sensitivity to the perceptual grouping of the patterned elements.

It is unclear whether successful pattern generalization was enabled because the perceptual features of motifs aligned with already known natural categories of song ele-

Fig. 4. Mean (±95% CI) normalized d' (methods), for subjects in the categorically-organized (green) and pseudo-randomly organized (purple) groups. *$p < 0.05$, see text for statistics. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ments i.e. warble and rattle, or because the categorical boundaries were more easily acquired during training due to greater within-class perceptual similarities, or both. Our choice to use of species-typical song elements was motivated by recent demonstrations that such stimuli, compared to simple tones, enhance task performance on other auditory tasks (Bregman, Patel, & Gentner, 2012). We suspect that starlings would perform similarly on this same pattern task using other auditory objects so long as they are sensitive to the perceptual categories of those stimuli – this remains, however, an open question for future research.

The present results may help in understanding mechanisms and constraints for pattern generalization in young infants. Specifically, Marcus, Fernandes, and Johnson (2007) report that infants fail to generalize patterning rules across sequences of tones, animal vocalizations or different musical timbres unless they are first exposed to those same patterning rules instantiated over speech sounds. Our results demonstrate a close interaction between pattern knowledge and pattern components, and support the idea that the abstraction of patterning information is closely constrained by the categorical structure of the constituent elements. Thus the ease with which children learn patterning rules instantiated over speech signals may reflect both familiarity and, perhaps more importantly, the perceptual structure of the speech sounds. This is consistent with several results suggesting that human infants use acoustic categories embedded in constituent elements of natural speech (e.g., words) as perceptual scaffolding for later acquiring knowledge of more abstract categories of language (Frigo & McDonald, 1998; Kelly, 1992; Monaghan, Christiansen, & Chater, 2007; Shi, Morgan, & Allopenna, 1998; Shi, Werker, & Morgan, 1999).

Our results indicate that the restricted focus of comparative language research on formal models of syntactic complexity is, at best, insufficient to understand how patterns are used. The abilities of non-human animals to perceive specific, formally defined patterns do not predict pattern generalization. Because reinterpreting patterns independent of specific perceptual events is essential for human language acquisition, we contend there is much to learn about the evolution of language faculties by understanding the benefits and constraints that learning any pattern exerts over multiple levels of abstraction.

Acknowledgements

References

Adret-Hausberger, M., & Jenkins, P. (1988). Complex organization of the warbling song in the European starling sturnus vulgaris. *Behaviour, 107*(3), 138–156.

Berwick, R. C., Beckers, G. J. L., Okanoya, K., & Bolhuis, J. J. (2012). A bird's eye view of human language evolution. *Frontiers in Evolutionary Neuroscience, 4*(April), 5. http://dx.doi.org/10.3389/fnevo.2012.00005.

Berwick, R., Okanoya, K., Beckers, G., & Bolhuis, J. (2011). Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences, 15*, 113–121.

Braaten, R. F. (2000). Multiple levels of representation of song by European starlings (*Sturnus vulgaris*): Open-ended categorization of starling sing types and differential forgetting of song categories and exemplars. *Journal of Comparative Psychology, 114*(1), 61–72. http://dx.doi.org/10.1037//0735-7036.114.1.61.

Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2012). Stimulus-dependent flexibility in non-human auditory pitch processing. *Cognition, 122*(1), 51–60. http://dx.doi.org/10.1016/j.cognition.2011.08.008.

Brown, P., & Jenkins, H. (1968). Auto-shaping of the pigeon's keypeck. *Journal of the experimental analysis of behavior, 11*(I), 1–8.

Eens, M., Pinxten, R., & Verheyen, R. F. (1988). Temporal and sequential organisation of song bouts in the starling. *Ardea, 77*, 75–86.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39*(2), 218–245. http://dx.doi.org/10.1006/jmla.1998.2569.

Gentner, T. Q. (2008). Temporal scales of auditory objects underlying birdsong vocal recognition. *The Journal of the Acoustical Society of America, 124*(2), 1350–1359.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*(7088), 1204–1207.

Gentner, T. Q., & Hulse, S. H. (1998). Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Animal Behaviour, 56*, 579–594.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99*(2), 349–364.

Kiggins, J. T., Comins, J. A., & Gentner, T. Q. (2012). Targets for a comparative neurobiology of language. *Frontiers in Evolutionary Neuroscience, 4*(April), 1–13. http://dx.doi.org/10.3389/fnevo.2012.00006.

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18*(5), 387–391. http://dx.doi.org/10.1111/j.1467-9280.2007.01910.x.

Margoliash, D., & Nusbaum, H. C. (2009). Language: the perspective from organismal biology. *Trends in Cognitive Sciences, 13*(12), 505–510.

Meliza, C. D. (2011). Effects of auditory recognition learning on the perception of vocal features in European starlings (*Sturnus vulgaris*). *The Journal of the Acoustical Society of America, 130*(5), 3115–3123. http://dx.doi.org/10.1121/1.3641420.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*(4), 259–305. http://dx.doi.org/10.1016/j.cogpsych.2006.12.001.

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language, 25*(1), 169–201.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*(2), B11–21.