

Research

Visual artificial grammar learning: comparative research on humans, kea (*Nestor notabilis*) and pigeons (*Columba livia*)

Nina Stobbe*, Gesche Westphal-Fitch, Ulrike Aust
and W. Tecumseh Fitch*

*Department of Cognitive Biology, Faculty of Life Sciences, University of Vienna, Althanstrasse 14,
1090 Vienna, Austria*

Artificial grammar learning (AGL) provides a useful tool for exploring rule learning strategies linked to general purpose pattern perception. To be able to directly compare performance of humans with other species with different memory capacities, we developed an AGL task in the visual domain. Presenting entire visual patterns simultaneously instead of sequentially minimizes the amount of required working memory. This approach allowed us to evaluate performance levels of two bird species, kea (*Nestor notabilis*) and pigeons (*Columba livia*), in direct comparison to human participants. After being trained to discriminate between two types of visual patterns generated by rules at different levels of computational complexity and presented on a computer screen, birds and humans received further training with a series of novel stimuli that followed the same rules, but differed in various visual features from the training stimuli. Most avian and all human subjects continued to perform well above chance during this initial generalization phase, suggesting that they were able to generalize learned rules to novel stimuli. However, detailed testing with stimuli that violated the intended rules regarding the exact number of stimulus elements indicates that neither bird species was able to successfully acquire the intended pattern rule. Our data suggest that, in contrast to humans, these birds were unable to master a simple rule above the finite-state level, even with simultaneous item presentation and despite intensive training.

Keywords: artificial grammar learning; formal language theory; pattern learning; visual stimuli; simultaneous presentation; comparative research

1. INTRODUCTION

The capacity to learn and recognize complex regularities and generalize over stimuli that follow abstract rules is thought to be a prerequisite for the ability to acquire language in humans [1–3]. One position concerning language evolution is that language acquisition mechanisms are part of a cognitive specialization for language acquisition and unique to humans [4,5]. An alternative perspective suggests that more general pattern-learning mechanisms are involved in language acquisition and use [6–8]. By this perspective, the general cognitive ability to extract regularities from patterns may constitute a domain-general learning mechanism shared with non-human animals ('animals', hereafter) [9]. Adjudicating among these possibilities (both of which may be partially correct) obviously requires direct

experimental comparison of the general pattern-processing abilities of humans and multiple non-human animal species. In this paper, we investigate the ability of humans and two bird species to recognize abstract, higher-order visual patterns, constructed using rules at two levels of computational complexity.

The artificial grammar learning (AGL) paradigm has been used widely since its first introduction in the 1960s [10,11] to explore the processes underlying rule acquisition processes. It typically involves the creation of an 'artificial grammar' involving one or more abstract rules. Subjects are exposed to stimuli conforming to these rules and subsequently tested for rejection of stimuli violating the rules. Additionally, researchers can test if subjects are able to generalize to novel stimuli, thus showing that they acquired a more abstract rule that is not bound to the original training stimuli. Despite the potentially misleading use of terms like 'grammar' and 'language' in this research area, the stimuli investigated have no meaning and this field has no direct connection to human language. These terms, which are borrowed from formal language theory, are used in a technical sense in which the term 'grammar' refers to some finite set of rules, and 'language' refers to

* Authors for correspondence (nina.stobbe@univie.ac.at; tecumseh.fitch@univie.ac.at).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rstb.2012.0096> or via <http://rstb.royalsocietypublishing.org>.

One contribution of 13 to a Theme Issue 'Pattern perception and computational complexity'.

string sets generated by such a grammar that include certain regularities [12].

Extensive AGL experiments with humans demonstrate that even infants and children are able to rapidly learn abstract rules underlying presented patterns without explicit instructions or feedback [13–16]. Based on the assumption that some of the learning mechanisms involved in AGL tasks might be shared with other animal species, AGL studies have also recently been carried out with different animal species [17–22], indicating that various animals are also able to recognize regularities in stimulus sets, and at least partly generalize over relatively simple sets of rules. However, the degree to which animals can process more computationally challenging rules, particularly those above the finite-state level, remains debated [11], and relatively few species have been tested so far.

Existing animal AGL findings thus suggest that comparative studies on a variety of species can yield further insight into the basic cognitive and neural mechanisms underlying rule learning and abstraction, and allow us to evaluate which such mechanisms are widely shared and which (if any) are unique to humans and/or language. So far, however, studies on animals have focused on one species each, and do not allow direct comparison of performance between different animal species. Comparative studies on different species that differ in biologically relevant traits (e.g. relative brain size or social complexity) may constitute a promising approach to investigate the evolution of rule learning abilities.

Because initial language learning in human infants is based on acoustic input, it is understandable that a variety of AGL studies in humans and nearly all AGL studies in animals have been carried out using acoustically presented stimuli. Single elements of acoustic stimuli (e.g. single syllables) are presented sequentially, thus requiring the mental storage of elements to obtain a complete acoustic representation of an entire pattern. Frank & Gibson [23] suggested that this memory demand can constitute a considerable constraint in rule learning tasks. When working memory load was reduced by presenting stimuli simultaneously, human participants were able to succeed in tasks in which they otherwise would have failed. This finding might be particularly relevant to comparative AGL experiments with humans and animals since the short-term memory capacity of different species may differ significantly.

Aiming to expand the current knowledge on comparative rule learning abilities in humans and animals, we designed a novel visual AGL task. Former studies on AGL by infants in different modalities indicate comparable learning mechanisms for auditory and visual stimuli that follow a predictable pattern [7]. However, while sequential learning performance is better in the auditory domain than in the visual domain, simultaneous presentation of visual stimuli leads to learning performance equal to sequential presentation of auditory stimuli [8,24–26]. Thus, to partially overcome possible memory constraints, we presented complex visual patterns in a simultaneous instead of a sequential manner. Additionally, we

aimed to avoid any learning bias due to potential relevance of stimuli such as human speech syllables or bird vocalizations. The visual stimuli were thus created from subunits that were non-representational tile images differing in many potential perceptual dimensions (including colour and shape), to rule out any influence of biological relevance of the stimuli to any species. These meaningless, abstract patterns were presented to humans as well as to two bird species, pigeons (*Columba livia*) and kea (*Nestor notabilis*).

Pigeons (*C. livia*) are birds with a relatively small brain size [27]. Their impressive performance on visual categorization tasks has been studied extensively, and they are readily able to learn arbitrary categories over images like the tiles making up our stimuli [28–31]. Kea (*N. notabilis*) are large brained parrots [27] and, as with most parrot species, lifelong vocal learners [32]. They are known for their playfulness and neophilia as well as for a variety of advanced cognitive abilities [33]. Neither the kea visual system nor kea's performance in visual tasks has been studied in detail. In our laboratories, both bird species readily work on touch-screen computers and are thus excellent subjects to compare a visual pattern learning task.

The first and central aim of this study was to find out whether all three species are able to discriminate between stimuli following two different pattern 'artificial grammars', with different levels of computational complexity in terms of formal language theory (a finite-state (AB)ⁿ grammar and a supra-regular AⁿBⁿ grammar). Second, we performed a crucial test for abstraction over the intended grammar, in which 'foil' stimuli were used that violated the grammar by including either one additional element or one element less than in the 'grammatical' stimuli. Correct rejection of these foil stimuli would require understanding of the underlying rule. Secondly, as a subsidiary goal, we explored our participants' generalization abilities by testing them with novel stimuli that differed from the original stimuli in various visual features. If subjects did not simply memorize specific training stimuli, they were expected to correctly choose 'grammatical' novel stimuli over 'non-grammatical' ones, despite superficial visual differences from the original stimuli. For these tests, we assessed whether subjects can generalize the learned rule to patterns of larger size than the training set, where the number of constituting elements is increased while the underlying rule stays the same. Humans' and birds' performance in this task can yield new insights into what types of regularities subjects are able to master in a learning task without explicit instructions.

2. METHODS

(a) *Stimuli*

Stimuli were abstract patterns made up of square tile-like elements ('tiles' hereafter) that belonged to two different and easily distinguishable categories. The single elements comprised 1 pixel black frames and internal complex geometrical patterns thereby constituting a visual analogue to the complex acoustic stimuli used in acoustic AGL studies. The tiles and the patterns were created algorithmically, using

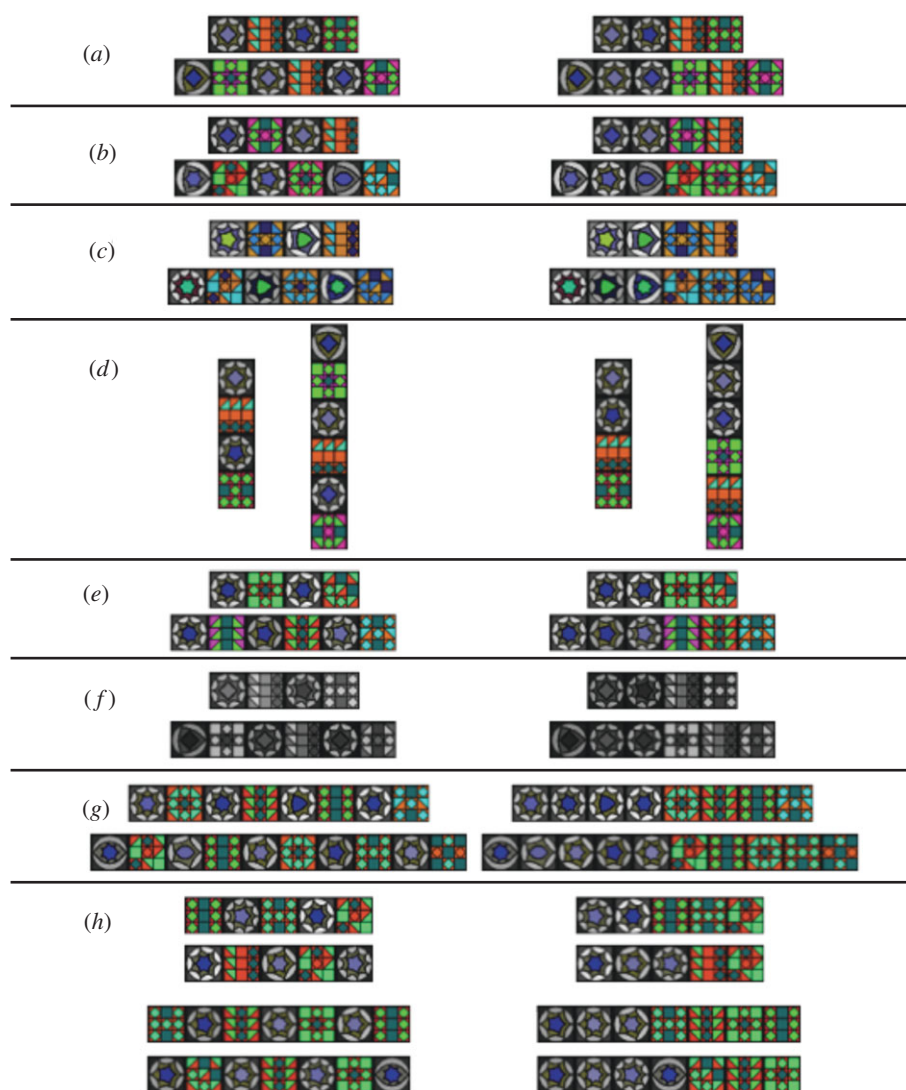


Figure 1. Stimuli used in initial training, generalization training and testing. Left column shows $(AB)^n$ stimuli, right column shows $A^n B^n$ stimuli. (a) Training stimuli, (b) novel stimuli, (c) stimuli with novel colours, (d) rotated stimuli, (e) scrambled stimuli, (f) greyscale stimuli, (g) extensions and (f) foil stimuli.

PYTHON (www.python.org) code implemented in NODEBOX (www.nodebox.net). One category of elements (A) included rounded, continuous shapes generated as Bezier curves and initial colours from the blue/grey spectrum. The other category (B) contained small angular polygons in initial colours from the red/green spectrum bordered by straight lines and were clearly distinguishable from the A-elements. We created 12 elements of each category that were then assembled according to two different rules (figure 1). Stimuli of the first grammar (hereafter referred to as $(AB)^n$) were made up of a string of AB units, i.e. A-elements and B-elements alternated, beginning with an A. Such patterns can easily be recognized by mechanisms at the lowest subregular level of computational complexity, the strictly local subset of the finite-state or regular grammars [12]. In the stimuli of the second grammar ($A^n B^n$), a group of A-elements was followed by a group of B-elements, where the number of A-elements matched the number of B-elements exactly in grammatical stimuli. The $A^n B^n$ language cannot be captured by a finite-state grammar and requires a context-free or stronger grammar [11]. One particular tile never appeared more than once in a stimulus. During initial

training, each specific A-element was arbitrarily paired with a specific B-element, forming constant A/B bigrams that were conserved but could fill any slot in the generated stimuli (e.g. for grammar 1: A1B1 A2B2 or A2B2 A1B1; for grammar 2: A1A2 B1B2 or A2A1 B2B1).

For the training phase, we created 240 patterned stimuli for each grammar ensuring that each stimulus of one grammar had a corresponding stimulus in the other grammar, containing the exact same elements. To expose subjects to a variety of grammatical patterns, half of the stimuli consisted of four elements (two As and two Bs) and half of them consisted of six elements (three As and three Bs; figure 1a).

After subjects reached criterion in this initial training, we embarked on a generalization training phase, designed to broaden their acceptance of patterns beyond the initial training stimuli, and incidentally to investigate which perceptual aspects of the initial stimuli were most salient. For this 'generalization' phase, we first ran five different test types, and continued to give feedback so that the participants continued to learn. Novel stimuli were all created according to the same rules as the training stimuli, and the simplest

set simply used novel arrangements of the identical tiles (*generalization test*; figure 1*b*). We also generated further stimuli using tiles that differed in the included colours. Colour ranges were shifted from blue/grey to green/gray in A-elements and from red/green to brown/blue for B-elements (*colour test*; figure 1*c*). To test whether orientation of the stimuli influences the ability of generalization, we also created stimuli that were rotated by 90° clockwise (*rotation test*; figure 1*d*). Since training stimuli comprised fixed bigrams of A- and B-elements, we scrambled all elements so that there was no longer a specific relationship between the elements (e.g. for grammar 1: A3B6 A5B1; *scrambled test*; figure 1*e*) in order to test generalization beyond the previously correlated bigrams. Finally, to further investigate the salience of colour cues, stimuli were additionally presented in greyscale, a transformation that removed all colour information for human subjects (*greyscale test*; figure 1*f*). Owing to differences in human and avian visual systems, birds most probably still perceived some colours in such stimuli [34], but the available colour information was also considerably reduced for the birds.

After completing this generalization training, we reached the crucial tests of the experiment: using unrewarded probe trials to determine if the intended 'grammar' had been mastered. To test for generalization beyond the n observed in training, stimuli with eight elements (four As and four Bs) and 10 elements (five As and five Bs) were created (*extensions*; figure 1*g*). Finally, unmatched *foil* stimuli were generated by either adding one element to a grammatical stimulus or removing one element, which resulted in unequal numbers of A- and B-elements (A^nB^m where $n \neq m$). This was done for stimuli previously consisting of two, three and four elements of each element type. Elements were removed or added either at the beginning of the stimulus or at the end leading to four different foil types per grammar type: $B(AB)^2$, $(AB)^2A$, $B(AB)^3$, $(AB)^3A$ and A^2B^3 , A^3B^2 , A^3B^4 , A^4B^3 , respectively (figure 1*h*).

(b) General procedure

For all species, subjects were divided into two groups for which either the $(AB)^n$ or the A^nB^n stimuli were the positive (rewarded) stimuli. To increase similarity to auditory AGL experiments, all subjects were initially presented with a short 'familiarization sequence' on the computer screen preceding each training session to prime the animals to the positive stimuli and facilitate learning. Familiarization sequences consisted of 30 stimuli (for kea) or 60 stimuli (for pigeons) of the positive grammar only (video sequence created with ALTERNATE PIC VIEW EXESLIDE). Each stimulus was presented for 900 ms, followed by a dark phase of 100 ms. Subjects were then trained and tested in a two alternative forced choice (2-AFC) procedure, and required to discriminate between stimuli from the two grammars. Each trial involved the simultaneous presentation of two stimuli in fixed positions on a black background on the computer screen. The left/right positions of the stimuli were randomized across trials. During training and generalization training (the first five experiments), subjects' choices were reinforced in all trials (see below for details) allowing learning.

During the crucial testing for generalization to *extensions* and *foil* stimuli, responses to probe trials were not reinforced. With one exception, positive stimuli were always presented simultaneously with the corresponding stimulus of the other grammar, i.e. a $(AB)^n$ stimulus was presented together with the corresponding A^nB^n stimulus made of the same tiles (thus preventing the use of any particular tile to discriminate between the patterns). However, in the *foil* test, to prevent the subjects from basing their decision purely on rejecting the obviously 'non-grammatical' stimuli, *foil* stimuli were presented together with the corresponding positive stimulus (with matched n). For an overview of methods for all three species, see the electronic supplementary material, table S1.

(c) Humans

Twenty human participants (14 females and six males) between 18 and 51 years old who had normal colour vision were tested in this study. Participants were not given any detailed information about the aim of the study before testing and instructions were reduced to the bare minimum needed, i.e. that they would see two images on the screen and would have to press either of two buttons to indicate their choice. All participants gave their written consent prior to participating and were paid €5 for their participation. To avoid distraction, participants were tested alone in a small room equipped with a computer and an IoLabs button box (www.iolab.co.uk). The experiment was run using custom code written in PYTHON. After watching the short familiarization sequence, people were instructed to wear headphones for acoustic feedback and start the training phase. During training, 30 trials were run in which subjects were asked to indicate their choice of one of the presented stimuli via a button press. Correct choices elicited a positive acoustic feedback tone (600 Hz, 0.5 s) and continuation to the next trial, incorrect choices led to a negative acoustic feedback sound (200 Hz, 0.5 s) and a red penalty screen for 3 s. Participants had to press a button within 5 s during training or 3 s during test, otherwise the image disappeared. These time limits were chosen in order to correspond to mean reaction times in birds. Participants were required to make at least 70 per cent first correct choices during training to proceed to the test phase; otherwise, they did not proceed to the test phase. In the generalization training block, 40 trials per test type (*generalization test*, *colour test*, *rotation test*, *scrambled test* and *greyscale test*) were presented in random order intermixed with 40 of the initial training stimuli, resulting in a test block of 240 trials. Acoustic and visual feedback was given after each trial. In the second test phase, 40 *extensions* and 80 *foils* were shown with no feedback given. After completion of the experiment, subjects were asked to describe their strategies in a short questionnaire. Only after providing their own opinion regarding what the task was about, were participants asked in more detail if they understood the rules involved, and if they had been counting elements and/or paying attention to symmetry. After concluding the experiments and debriefing, each participant was given detailed information about the study aims.

(d) Birds

Twelve kea (*N. notabilis*) and 10 pigeons (*C. livia*) participated in the study. Owing to the death of some subject birds during the experiments, sample sizes vary between tests. Kea were housed in a group of 21 individuals in a large outdoor aviary (about 520 m²) at the Haidlhof Research Station, Bad Vöslau, and were fed three times a day. Pigeons were housed in outdoor aviaries at the University of Vienna in groups of about eight individuals and were maintained at slightly below (about 90% of) their free-feeding weight. All birds were familiar with a touch screen and the general procedure of a two-choice task but naïve to this specific task before beginning training. Kea were trained individually in an experimental chamber which was open at one side, allowing the birds to enter voluntarily. Pigeons were individually placed in separate, closed indoor Skinner boxes by the experimenter. Birds indicated their choice by pecking on a 15 inch TFT computer screen mounted behind an infrared touch frame (Carroll Touch, 15"). Food reward was dispensed by means of a special feeder that released a portion of peanut (1/8) for kea, or a small amount of grain for pigeons, to a small food repository directly below the touch screen. Data acquisition and device control were handled with hardware and software especially developed for the requirements of various cognitive experiments (COGNITIONLABLIGHT, v. 1.9, © M. Steurer).

During presentation of the familiarization sequences, birds were prevented from interacting with the screen. For pigeons a Plexiglas barrier was placed between the bird and the screen; images were enlarged to equalize for greater viewing distance. For kea, the touch function of the screen was disabled during the presentation. Before starting the training session, the Plexiglas barrier was removed and touch function was enabled, respectively, so birds could indicate their choice by pecking on one of the presented stimuli. Correct choices led to disappearance of the images and were reinforced by a positive acoustic feedback tone (600 Hz, 0.5 s) and food reward. A peck on an incorrect stimulus caused a correction trial with a negative feedback sound (200 Hz, 0.5 s). The screen turned red for 3 s, after which the same pair of stimuli was presented again. This continued until the bird pecked the correct stimulus. Each trial was followed by a 4 s intertrial interval during which the screen was black. Birds normally completed a session including 40 trials per day, aborted sessions were continued at the point of stopping, the following day.

Training was terminated when birds had completed a certain minimum number of sessions (18 sessions for kea, 24 sessions for pigeons) and performance had fulfilled a pre-specified learning criterion. This was set to at least 70 per cent first correct choices per session in six consecutive sessions (corresponding to $p < 0.008$ in a one-sided binomial test). Pigeons were allowed to repeat one session if five out of six sessions were significant. If the repeated session was significant, the criterion was considered to be satisfied.

Subsequent generalization training (*generalization test*, *colour test*, *rotation test*, *scrambled test* and *greyscale test*) included 20 test trials per session, randomly intermixed with 20 of the original training trials. For each

of these tests, birds had to complete 12 sessions in which test trials were reinforced like training trials. Reinforcement was maintained to avoid a failure in transfer performance owing to neophobia, a well-known influencing variable in transfer tests with pigeons [35,36]. This procedure allows further learning after initial training and at the same time enlarges the variety of stimulus types the subjects were confronted with before proceeding to the crucial non-reinforced 'probe' testing. The general operant contingencies were in accordance with other studies testing pattern learning in birds [20,22]. As most birds had severe difficulties with the greyscale stimuli, all birds received additional training that was identical to the initial training except that the stimuli were replaced with their greyscale versions and preceded by greyscale familiarization sequences. Greyscale training was ended when either the same criterion was reached as in the initial training or when a bird had completed as many sessions as the slowest bird in the initial training. After completing the greyscale training, birds were subjected to another greyscale test with greyscale versions of novel stimuli.

Probing for generalization to *extensions* and *foils* was non-reinforced, i.e. the first peck on either of the two presented stimuli terminated the trial without food reward, acoustic feedback or correction trial. To avoid frustration in the birds, we reduced the number of probe trials per session to eight trials out of 40. Probing for generalization to *extensions* comprised 80 probe trials shown across 10 sessions. For the *foils*, probe trials in each session were embedded in a set of 16 training trials and eight trials presenting two correct stimuli simultaneously ('double S+'). Since *foil* stimuli were not presented with a 'non-grammatical' stimulus as in all tests before but with a 'grammatical' stimulus, 'double S+' trials were included, to prevent birds from recognizing non-reinforced trials that simply lack a stimulus of the other grammar. Eighty stimulus pairs were presented per *foil* type (see above) leading to a total of 40 test sessions.

(e) Statistical analysis

Statistical significance in training and test phase performance was analysed with a one-sided binomial test including number of correct first choices, total number of trials and a confidence level of 0.95. For each species, a generalized estimating equations model for repeated measures was fitted to the data, taking account of grammar type and test type as factors. Further comparisons between grammar types and test types were carried out with post hoc comparisons including Bonferroni corrections. Statistical analysis was performed with R v. 2.12.0 and PASW Statistics v. 18.0.

3. RESULTS**(a) Humans**

All 20 participants reached the training criterion of at least 70 per cent first correct choices within a set of 30 training trials. In the following test trials, performance did not differ between the two groups trained on either (AB)ⁿ or AⁿBⁿ (factor group: Wald $\chi^2 = 3.43$,

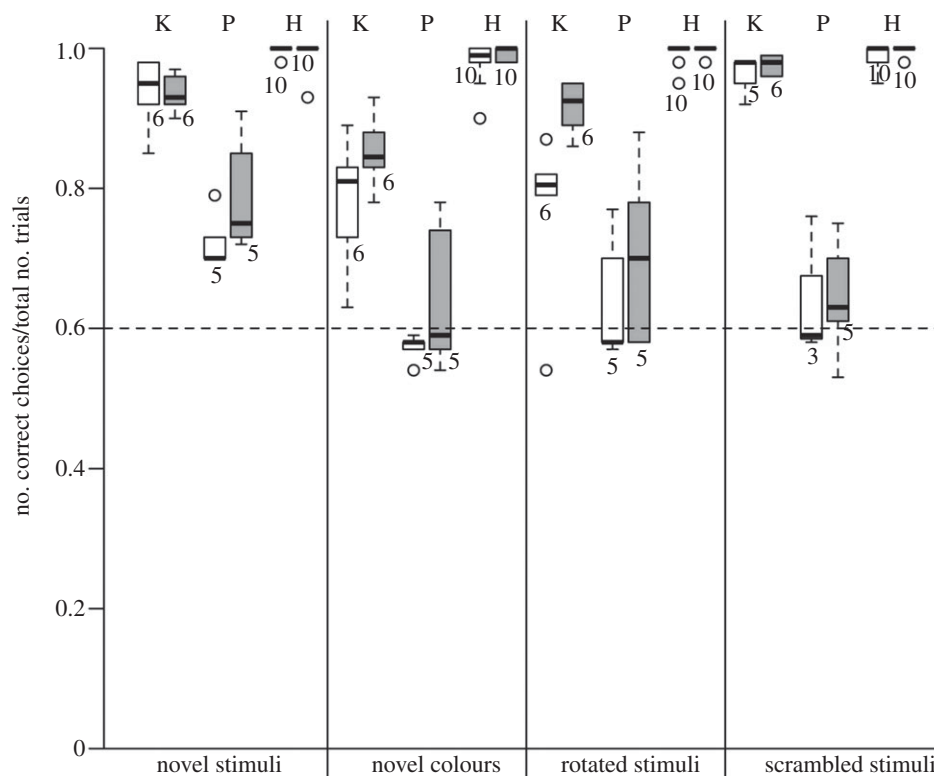


Figure 2. Performance of kea (K), pigeons (P) and humans (H) in four different generalization training tasks. White symbols indicate subjects trained on $(AB)^n$ grammar, grey symbols indicate subjects trained on A^nB^n grammar. Numbers below boxplots specify sample size. The dashed line marks the level of significance (binomial test).

$p = 0.06$). Performance in the generalization tasks (*generalization test*, *colour test*, *rotation test*, *scrambled test* and *greyscale test*) was at a very high level: subjects of both groups chose the correct stimulus in more than 90 per cent of the 240 trials (figures 2 and 3). In the crucial unrewarded test trials, humans generalized the pattern rule to stimuli with extended numbers of tiles ($n = 4$ and 5 ; figure 4) without difficulty. Performance with unmatched foil stimuli, with either an element added or taken away, varied between the two grammars. Although performance significantly dropped for both (factor test: Wald $\chi^2 = 43.27$, $p > 0.001$; figure 4), only six out of 10 participants trained on $(AB)^n$ performed significantly above chance, while eight out of 10 participants in the A^nB^n group passed the test, scoring more than 70 per cent correct first choices (see the electronic supplementary material, table S2). These results indicate that, while most human participants correctly grasped the regularity underlying the supra-regular A^nB^n grammar, nearly half of the participants did not acquire the intended rule with the $(AB)^n$ grammar.

In the $(AB)^n$ group, successful participants reported that they based their decision only on the first and last elements of the stimuli that had to be different (A-element at the start, B-element at the end) to match the learned pattern rule. Only one participant reported counting the number of elements and choosing the stimuli with equal numbers of A- and B-elements. In the A^nB^n group, the eight successful participants stated that they made their choice based on the visual perception of symmetry arising from equal numbers of A-elements on the left and

B-elements on the right side of the stimulus ($n = 2$), by counting the elements ($n = 2$), or by some combination of the two strategies ($n = 4$).

(b) Birds

Both bird species successfully learned to discriminate between $(AB)^n$ and A^nB^n stimuli. While kea in both groups passed the training phase on average in 15 ± 1 sessions (mean \pm s.e.; corresponding to 600 ± 40 trials), most pigeons took much longer to reach the learning criterion ($(AB)^n$ group: 85 ± 20 sessions, corresponding to 3400 ± 800 trials; A^nB^n group: 53 ± 24 sessions, corresponding to 2120 ± 960 trials). In subsequent phases, pigeons did not show differences between the two grammars (Wald $\chi^2 = 0.87$, $p = 0.35$). In kea, however, birds trained to peck on A^nB^n stimuli performed better overall than birds in the $(AB)^n$ group (Wald $\chi^2 = 6.65$, $p = 0.01$).

During generalization training, the type of test had a significant influence on performance in both species (Wald $\chi^2 = 10253.66$ and 2514.86 , both $p < 0.001$). When confronted with the first four generalization tasks, the level of performance was generally higher for kea than for pigeons (figure 2). All kea successfully generalized the pattern rule to novel stimuli, novel colours, rotated and scrambled stimuli. Pigeons readily generalized to novel stimuli using the same tiles as the training stimuli, but many of the pigeons did not pass the significance level with the other three types of stimuli (see electronic supplementary material, table S2).

The considerable reduction of colour information by presenting stimuli in greyscale led to severe

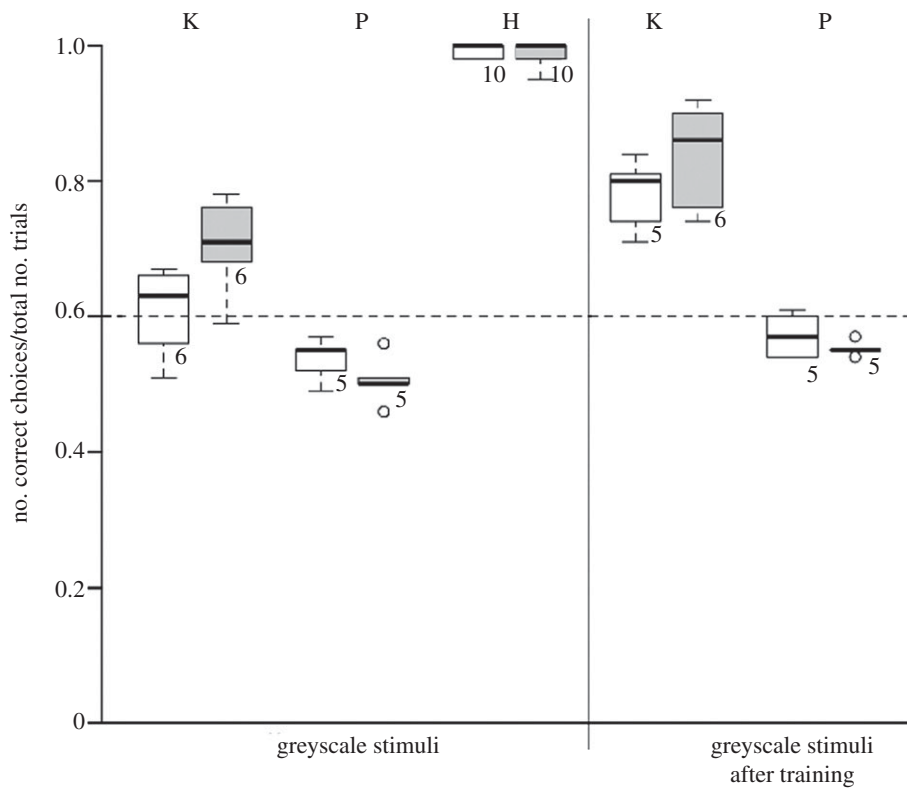


Figure 3. Performance of kea (K), pigeons (P) and humans (H) in generalization tasks with greyscale stimuli before and after additional training. White symbols indicate subjects trained on $(AB)^n$ grammar, grey symbols indicate subjects trained on A^nB^n grammar. Numbers below boxplots specify sample size. The dashed line marks the level of significance (binomial test).

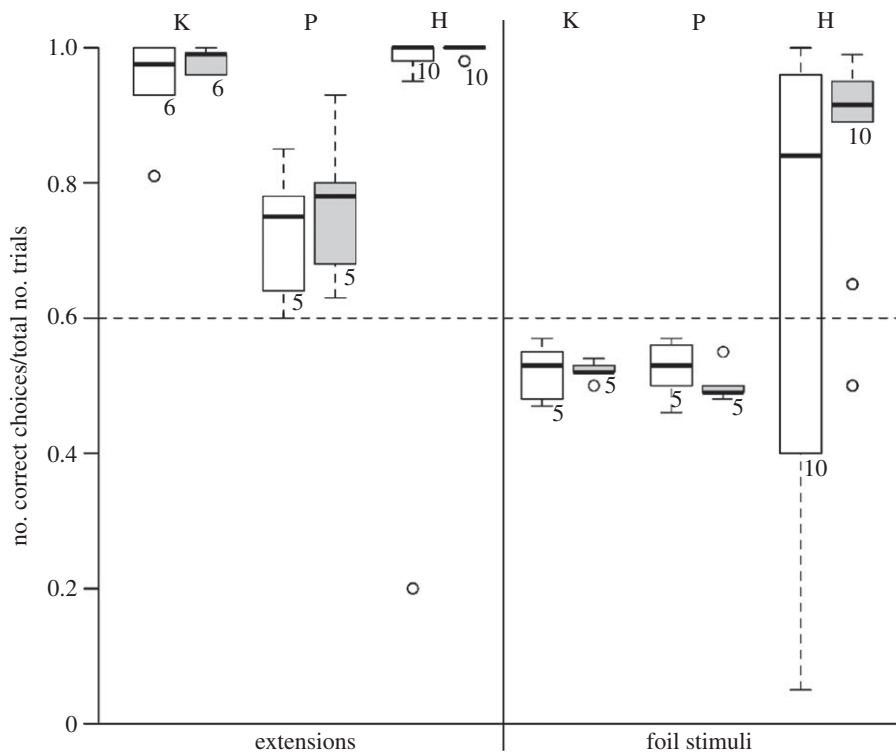


Figure 4. Performance of kea (K), pigeons (P) and humans (H) in tasks with extended stimuli ($n = 4$ and 5) and mismatched foil stimuli. White symbols indicate subjects trained on $(AB)^n$ grammar, grey symbols indicate subjects trained on A^nB^n grammar. Numbers below boxplots specify sample size. The dashed line marks the level of significance (binomial test).

decrement of performance in both bird species compared with their performance in the first generalization task (both $p < 0.001$). While kea still

chose the correct stimulus more often than predicted by chance, pigeons of both groups were no longer able to differentiate between $(AB)^n$ and A^nB^n stimuli

(figure 3). Subsequent training of all birds with grey-scale stimuli led to successful learning only in kea ((AB)ⁿ group: 14 ± 5 sessions to achieve six consecutive significant sessions, AⁿBⁿ group: 10 ± 2 sessions); pigeons remained at chance level even after receiving 159 rewarded training sessions (corresponding to the number of sessions needed by the slowest bird in initial training). Retesting the birds with novel greyscale stimuli after the additional training revealed significantly improved performance in kea (both grammars, $p < 0.001$), but no improvement in pigeons ($p = 0.99$). In short, pigeons failed completely to discriminate among stimuli when colour cues were reduced. In summary, in the generalization training phase, kea successfully generalized to new shapes, colours, orders and orientations, and with training, to stimuli mostly lacking colour. In contrast, pigeons had difficulty with all generalizations, and even with prolonged intensive training were unable to cope with the grey-scale stimuli.

In the final crucial unrewarded probe tests, we investigated what, precisely, the birds had learned. *Extension* probes with either four or five elements per tile type did not impair performance compared with the first generalization task in either bird species (both $p > 0.45$). However, both species failed entirely on the mismatched *foil* probes: when birds had to choose between a correct, matched stimulus and a foil stimulus that deviated from the intended grammar either owing to an additional element or owing to one element less, members of both species chose randomly without any preference for either stimulus type (figure 4). We conclude from this that, despite their various successful generalizations, neither bird species acquired the intended grammar. In particular, we found no evidence that either kea or pigeons correctly induced the supra-regular AⁿBⁿ grammar.

For the majority of participants in all three species, error rates did not differ depending on where in the foil stimulus an element was added or removed (one-sided binomial tests, all $p > 0.05$). However, one kea and one human subject of the (AB)ⁿ group showed a certain pattern in their discrimination errors. The kea made significantly more errors when the foil stimuli started with a B-element than when the stimuli ended with an A-element ($p = 0.01$). The human subject, on the other hand, showed the reversed pattern, making significantly more errors with stimuli ending with an A-element ($p = 0.01$).

4. DISCUSSION

Our results clearly show that all participants were able to discriminate between two high-level patterns: training of both humans and two bird species in a 2-AFC resulted in successful discrimination of two types of visual patterns, each structured according to different rules. This result is consistent with many previous studies showing that humans share basic visual pattern recognition abilities with other animals, even with insect species [37,38]. Human subjects rapidly acquired the task within the first 30 trials of reinforced learning. Both bird species required a much larger number of training sessions to reliably choose the

correct stimulus; while the slowest kea reached the learning criterion after 1000 trials, the slowest pigeon took over 6000 trials to perform above chance level. Pigeons' learning performance in this task, however, was still faster than learning performance in a previous AGL task involving coloured letters [39] or acquisition time for starlings learning to discriminate between auditory stimuli following the same patterns [20]. General levels of performance were considerably lower for pigeons than for the other two species, but comparable to previous AGL studies in this species [39,40].

Both humans and kea were able to generalize beyond the training stimuli: subsequent transfer to novel stimuli was successful when the stimuli consisted of the same tiles arranged in different orders, new tiles with novel colours, when the whole pattern was rotated around 90° counterclockwise or consisted of scrambled A- and B-elements (eliminating A/B-correspondences present in the training stimuli). The ability to transfer pattern discrimination to novel instances suggests that with training both species can generalize beyond specific features of the training stimuli such as single elements, A/B bigrams or colour configuration, to acquire a more general pattern rule.

In contrast, pigeons only performed highly above chance level in the first generalization task, using identical tile elements in new orders, and many pigeons showed difficulties in applying the pattern rule to novel stimuli with changed visual features, suggesting that pigeons tend to rely on more stimulus-specific features in the initial discrimination acquisition. This assumption is consistent with former studies on pigeons' visual discrimination learning indicating that pigeons tend to focus on the most salient visual cue. If available they mostly respond consistently to the colour dimension in visual discrimination tasks [41–43], and often fail when presented with complex problems that require the application of a more abstract rule [44]. Surprisingly, some pigeons trained on (AB)ⁿ stimuli also showed near chance-level performance when stimuli were rotated. This outcome stands in contrast to the earlier findings suggesting that pigeons show rotational invariance [45].

Both bird species exhibited a significant drop of performance when available colour information was drastically reduced (*greyscale test*). This phenomenon is consistent with previous studies on visual discrimination [42,46,47]. Pigeons' inability to discriminate between the two types of visual patterns in greyscale even after extensive training further supports the assumption that they based their prior decisions primarily on available colour cues. Human subjects, on the other hand, continued to perform at high levels even when colour cues were removed, suggesting that they inferred a pattern rule that was independent of colour configurations. Some kea also seemed to base their decisions strongly on colour cues, but were able, after additional training, to correctly respond to greyscale stimuli. Test trials during the above generalization training phase were reinforced, rewarding correct choices and triggering correction trials in the case of incorrect choices, so subjects still had the opportunity to learn from their errors (see electronic supplementary material, figure S1).

The overall better performance of the A^nB^n group compared with the $(AB)^n$ group in kea is somewhat surprising, given that a grammar with alternating 'A's and 'B's is classified as a finite-state grammar and thought to be easier to learn than a supra-regular grammar [see 11,12]. This counterintuitive result might be explained by the specific setup of our experiments. Clusters of 'A's and 'B's might have been easier and faster to recognize from a perceptual point of view than stimuli with alternating elements that are perceptually more complex. This hypothesis, however, remains speculative at present, because we do not know whether subjects actively chose the 'grammatical' stimulus or actively rejected the 'non-grammatical' one. Future studies in which stimulus elements are presented sequentially rather than simultaneously might provide deeper insight into perceptual strategies that underlie complex pattern learning.

We now turn to the crucial last two experiments which used unrewarded probe trials. Success on the *Extension* test clearly demonstrates the application of some type of pattern rule, independent of feedback, in all three species. All three species correctly classified stimuli with two or four ($n = 4$ and 5 , respectively) additional elements as belonging to the learned class of correct stimuli, successfully generalizing over stimulus length. Subjects could have passed all of the tests so far by perceptually discriminating between stimuli consisting of alternating 'A's and 'B's versus clusters of 'A's followed by clusters of 'B's without matching the numbers of elements. In the final *foil* test, we presented the subjects with a choice between a correct stimulus and a stimulus in which one element was added or removed. Neither kea nor pigeons rejected such unmatched foil stimuli, choosing between correct and foil stimuli in a random manner. We conclude therefore that the birds did not successfully acquire the intended grammars. This is particularly relevant for the supra-regular grammar A^nB^n , because it is precisely the match between the two components of the pattern which requires a context-free grammar or higher. In contrast, eight of 10 human participants spontaneously rejected such mismatched foils in this grammar (in contrast to [48]).

We found that birds were able to achieve a high level of success on various types of generalization, but nonetheless did not reject key violations of the intended grammar. This failure clearly illustrates the need for a thorough, by-category analysis of AGL results and a careful assessment of their implications (cf. [22,49–51]). Many of our generalization tests can be solved based on alternative strategies (e.g. clusters versus alternations of elements) that do not correspond to the intended 'grammar'. The use of alternative strategies is to be expected when multiple generalizations are possible based on the initial stimuli [52,53]. Alternative strategies can lead to above chance performance in a variety of generalization tasks, potentially leading to a false conclusion that subjects have acquired the precise grammar intended by the experimenter [54]. Based on this finding, it is also important to analyse individual performance instead of grouping subjects to be able to pinpoint individual learning strategies [22,54].

In our study, subjects that failed to reject foil stimuli must have acquired alternative strategies allowing

them to choose the correct stimulus well above chance level in the generalization tasks that required transfer to novel stimuli. Given that the stimuli presented during training and first generalization tests did not force the subjects to pay attention to the matching numbers of 'A's and 'B's, this feature was not always included in the acquired rule even in humans, and was never included by either bird species [52,53,55]. Moreover, simultaneous presentation of 'grammatical' and 'non-grammatical' stimuli throughout the main parts of our study might have led subjects to base their decisions, and thus their acquired rule, mainly on the differences between the two stimulus classes. Detailed analysis of individual performance revealed that for the vast majority of subjects, the position of the incorrect element did not influence performance. However, one kea of the $(AB)^n$ group was significantly better in performance in trials where foil stimuli illegitimately ended with an A-element, suggesting that this bird applied a recency rule [22,49], focused on the last elements of the stimulus, and one human participant of the $(AB)^n$ group showed the opposite pattern, significantly more often rejecting the foil stimuli that started with a B-element (i.e. applying a primacy rule). Further ongoing experiments including 'non-grammatical' stimuli that vary in the level of similarity to the 'grammatical' stimuli and a more detailed analysis of individual performance will allow a better understanding of what exact strategies were applied by the subjects.

Surprisingly, although performance in all previous tasks was near the ceiling level, even some of our human participants failed to reject mismatched foils. On the one hand, in the A^nB^n group, eight of 10 human participants successfully chose correct 'matched' stimuli over the foil stimuli. These participants explicitly reported that they primarily used symmetry features and/or counting [22,54]. In contrast, in the $(AB)^n$ group four of 10 happily accepted mis-matched strings, indicating that they did not acquire the intended strictly local grammar. The six of 10 rejecting mismatched foils reported comparing the first and last elements, which had to be different. Interestingly, this depends on a long-distance relationship in the pattern.

Either counting or symmetry-based strategies appear to pose major challenges for our two tested bird species. Although many bird species have shown the ability to form a concept of numerosity, it remains unclear if this ability is primarily based on conceptual subitizing or actually on counting elements and up to which numbers this ability can go in birds [27,56]. Furthermore, counting alone is inadequate to solve the A^nB^n task: two counts must be made, of 'A's and 'B's, and then compared. Both pigeons and starlings show difficulties in learning discriminations based upon symmetry [57,58]. We suggest that our avian subjects failed the foil tests because they acquired a pattern rule based on alternating versus block-consistent structure of A- and B-elements (i.e. they detected local dependencies, capturable by a finite-state grammar), but ignored the total numbers of 'A's and 'B's [11,12]. However, the ability of kea to generalize to 'scrambled' tiles, in which the specific A/B correspondences were broken, suggests that they did not simply memorize

bigrams in order to recognize the patterns, as has been suggested for humans ([49], but see also [59]).

Our study represents the first attempt to study mechanisms of AGL across three different species. The results show that both bird species failed to learn the intended grammatical rule, but nonetheless developed alternative strategies enabling them to solve the generalization tasks to a considerable extent. We assume that 'configural processing' *sensu* Maurer et al. [60] was involved in learning mechanisms of humans and kea, leading to the extraction of certain relations among features of the compound stimuli, allowing them to apply the learned rule to a variety of novel stimuli. The pigeon data in contrast suggest that these birds dominantly apply a form of 'featural processing' that takes into account only single features but not general relationships among them. Since pigeons are the species with the smallest relative brain size within our three study species, our data support the hypothesis that relative brain size may be a factor that considerably influences the ability to detect more complex relationships between pattern elements [61].

In summary, our human results confirm the ability of humans to acquire abstract visual patterns generated by a simple supra-regular grammar A^nB^n , both extending to novel n and rejecting mismatched ns . These findings support the hypothesis that human pattern-processing capabilities are not limited to patterns made up of linguistic items (auditory syllables or written letters) but readily extend to abstract, meaningless visual images. In contrast, our bird results suggest that neither kea nor pigeons deduce such a supra-regular grammar, but instead make use of various lower-level rules to discriminate among such patterns. These results contrast with findings using auditory AGL in starlings ([20], but see also [22]) and Bengalese finches ([62], but see also [63]), but are consistent with other animal results [17,22]. Our results provide no evidence that complex pattern perception abilities, used across sensory domains by humans, are shared with these other species. More generally, the new visual AGL paradigm we introduce here provides a relatively level playing field for comparative tests among different animal species, and clearly highlights cognitive advantages of kea over pigeons [33]. Finally, our visual paradigm can easily be extended to sequential visual presentation, to allow precise titration of working memory demands during processing of otherwise identically patterned visual and auditory stimuli. We thus suggest that it will provide a useful addition to the empirical toolkit used in AGL research to compare abstract pattern perception across multiple species.

Animal housing and experimental setup followed the Animal Behavior Society Guidelines for the Use of Animals in Research, the legal requirements of Austria and all institutional guidelines.

We gratefully acknowledge the help of J. and K. Kramer and M. Schlumpp in data collection. This research was funded by an ERC Advanced Grant SOMACCA to WTF.

REFERENCES

- 1 Brown, R. 1973 *A first language: the early stages*. Cambridge, MA: Harvard University Press.
- 2 Pinker, S. 1994 *The language instinct*. New York, NY: Morrow.
- 3 Cook, R. G. & Brooks, D. I. 2009 Generalised auditory same-different discrimination by pigeons. *J. Exp. Psychol. Anim. Behav. Process.* **35**, 108–115. (doi:10.1037/a0012621)
- 4 Chomsky, N. 1957 *Syntactic structures*. The Hague, The Netherlands: Mouton.
- 5 Jackendoff, R. 2002 *Foundations of language: brain, meaning, grammar, evolution*. New York, NY: Oxford University Press.
- 6 Christiansen, M. H. & Chater, N. 2008 Language as shaped by the brain. *Behav. Brain Sci.* **31**, 536–537. (doi:10.1017/S0140525X0800527X)
- 7 Kirkham, N. Z., Slemmer, J. A. & Johnson, S. P. 2002 Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* **83**, B35–B42. (doi:10.1016/S0010-0277(02)00004-5)
- 8 Saffran, J. R., Pollak, S. D., Seibel, R. L. & Shkolnik, A. 2007 Dog is a dog is a dog: infant rule learning is not specific to language. *Cognition* **105**, 669–680. (doi:10.1016/j.cognition.2006.11.004)
- 9 Endress, A. D., Cahill, D., Block, S., Watumull, J. & Hauser, M. D. 2009 Evidence for an evolutionary precursor to human language affixation in a non-human primate. *Biol. Lett.* **5**, 749–751. (doi:10.1098/rsbl.2009.0445)
- 10 Reber, A. S. 1967 Implicit learning of artificial grammars. *J. Verb. Learn. Verb. Behav.* **6**, 855–863. (doi:10.1016/S0022-5371(67)80149-X)
- 11 Fitch, W. T. & Friederici, A. D. 2012 Artificial grammar learning meets formal language theory: an overview. *Phil. Trans. R. Soc. B* **367**, 1933–1955. (doi:10.1098/rstb.2012.0103)
- 12 Jäger, G. & Rogers, J. 2012 Formal language theory: refining the Chomsky hierarchy. *Phil. Trans. R. Soc. B* **367**, 1956–1970. (doi:10.1098/rstb.2012.0077)
- 13 Aslin, R. N., Saffran, J. R. & Newport, E. L. 1998 Computation of conditional probability statistics by 8-month-old infants. *Psychol. Sci.* **9**, 321–324. (doi:10.1111/1467-9280.00063)
- 14 Gomez, R. L. & Gerken, L. 1999 Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* **70**, 109–135. (doi:10.1016/S0010-0277(99)00003-7)
- 15 Marcus, G. F., Vijayan, S., Rao, S. B. & Vishton, P. M. 1999 Rule learning by seven-month-old infants. *Science* **283**, 77–80. (doi:10.1126/science.283.5398.77)
- 16 White, K. S., Peperkamp, S., Kirk, C. & Morgan, J. L. 2008 Rapid acquisition of phonological alternations by infants. *Cognition* **107**, 238–265. (doi:10.1016/j.cognition.2007.11.012)
- 17 Fitch, W. T. & Hauser, M. D. 2004 Computational constraints on syntactic processing in a nonhuman primate. *Science* **303**, 377–380. (doi:10.1126/science.1089401)
- 18 Newport, E. L., Hauser, M. D., Spaepen, G. & Aslin, R. N. 2004 Learning at a distance. II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cogn. Psychol.* **49**, 85–117. (doi:10.1016/j.cogpsych.2003.12.002)
- 19 Toro, J. M. & Trobalón, J. B. 2005 Statistical computation over a speech stream in a rodent. *Atten. Percept. Psychophys.* **67**, 867–875. (doi:10.3758/BF03193539)
- 20 Gentner, T. Q., Fenn, K. M., Margoliash, D. & Nusbaum, H. C. 2006 Recursive syntactic pattern learning by songbirds. *Nature* **440**, 1204–1207. (doi:10.1038/nature04675)
- 21 Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F. & Cushman, F. 2007 Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* **107**, 479–500. (doi:10.1016/j.cognition.2007.10.010)

- 22 Van Heijningen, C. A. A., de Visser, J., Zuidema, W. & ten Cate, C. 2009 Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl Acad. Sci. USA* **106**, 20 538–20 543. (doi:10.1073/pnas.0908113106)
- 23 Frank, M. C. & Gibson, E. 2011 Overcoming memory limitations in rule learning. *Lang. Learn. Dev.* **7**, 130–148. (doi:10.1080/15475441.2010.512522)
- 24 Hunt, R. H. & Aslin, R. N. 2001 Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *J. Exp. Psychol.* **130**, 658–680. (doi:10.1037/0096-3445.130.4.658)
- 25 Fiser, J. & Aslin, R. N. 2002 Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 458–467. (doi:10.1037/0278-7393.28.3.458)
- 26 Saffran, J. R. 2002 Constraints on statistical language learning. *J. Mem. Lang.* **47**, 172–196. (doi:10.1006/jmla.2001.2839)
- 27 Emery, N. J. 2006 Cognitive ornithology: the evolution of avian intelligence. *Phil. Trans. R. Soc. B* **361**, 23–43. (doi:10.1098/rstb.2005.1736)
- 28 Aust, U. & Huber, L. 2002 Target-defining features in a ‘people-present/people-absent’ discrimination task by pigeons. *Learn. Behav.* **30**, 165–176. (doi:10.3758/BF03192918)
- 29 Lazareva, O. F., Freiburger, K. L. & Wasserman, E. A. 2004 Pigeons concurrently categorize photographs at both basic and superordinate levels. *Psychon. Bull. Rev.* **11**, 1111–1117. (doi:10.3758/BF03196745)
- 30 Aust, U., Apfalter, W. & Huber, L. 2005 Pigeon categorization: classification strategies in a non-linguistic species. In *Images and reasoning. Interdisciplinary conference series on reasoning studies, 1* (eds P. Grialou, G. Longo & M. Okada), pp. 183–204. Tokyo, Japan: Keio University Press.
- 31 Huber, L. & Aust, U. 2006 A modified feature theory as an account to pigeon visual categorization. In: Categories and concepts in animals, T. Lazareva, O. F., & Wasserman, E. A. (eds). In *Learning theory and behavior, vol. 1. Learning and memory: a comprehensive reference* (ed. J. Byrne), pp. 197–226. Oxford, UK: Elsevier.
- 32 Bond, A. B. & Diamond, J. 2005 Geographic and ontogenetic variation in the contact calls of the kea (*Nestor notabilis*). *Behaviour* **142**, 1–20. (doi:10.1163/1568539053627721)
- 33 Huber, L. & Gajdon, G. K. 2006 Technical intelligence in animals: the kea model. *Anim. Cogn.* **9**, 295–305. (doi:10.1007/s10071-006-0033-8)
- 34 Cuthill, I. C., Hart, N. S., Partridge, J. C., Bennett, A. T. D., Hunt, S. & Church, S. C. 2000 Avian colour vision and avian video playback experiments. *Acta Ethol.* **3**, 29–37. (doi:10.1007/s102110000027)
- 35 Clement, T. S. & Zentall, T. R. 2003 Choice based on exclusion in pigeons. *Psychon. Bull. Rev.* **10**, 959–964. (doi:10.3758/BF03196558)
- 36 Blaisdell, A. P. & Cook, R. G. 2005 Two-item same-different concept learning in pigeons. *Learn. Behav.* **33**, 67–77. (doi:10.3758/BF03196051)
- 37 Heisenberg, M. 1995 Pattern recognition in insects. *Curr. Opin. Neurobiol.* **5**, 475–481. (doi:10.1016/0959-4388(95)80008-5)
- 38 Avarguès-Weber, A., Portelli, G., Benard, J., Dyer, A. & Giurfa, M. 2009 Configural processing enables discrimination and categorization of face-like stimuli in honeybees. *J. Exp. Biol.* **213**, 593–601. (doi:10.1242/jeb.039263)
- 39 Herbranson, W. T. & Shimp, C. P. 2008 Artificial grammar learning in pigeons. *Learn. Behav.* **36**, 116–137. (doi:10.3758/LB.36.2.116)
- 40 Herbranson, W. T. & Shimp, C. P. 2003 ‘Artificial grammar learning’ in pigeons: a preliminary analysis. *Learn. Behav.* **31**, 98–106. (doi:10.3758/BF03195973)
- 41 Jones, L. V. 1954 Distinctiveness of color, form and position cues for pigeons. *J. Comp. Physiol. Psychol.* **47**, 253–257. (doi:10.1037/h0058499)
- 42 Huber, L., Troje, N. F., Loidolt, M., Aust, U. & Grass, D. 2000 Natural categorization through multiple feature learning in pigeons. *Q. J. Exp. Psychol.* **53B**, 341–357. (doi:10.1080/027249900750001347)
- 43 Kirsch, J. A., Kabanova, A. & Güntürkün, O. 2008 Grouping of artificial objects in pigeons: an inquiry into the cognitive architecture of an avian mind. *Brain Res. Bull.* **75**, 485–490. (doi:10.1016/j.brainresbull.2007.10.033)
- 44 Aust, U., Range, F., Steurer, M. & Huber, L. 2008 Inferential reasoning by exclusion in pigeons, dogs, and humans. *Anim. Cogn.* **11**, 587–597. (doi:10.1007/s10071-008-0149-0)
- 45 Hollard, V. D. & Delius, J. D. 1982 Rotational invariance in visual pattern recognition by pigeons and humans. *Science* **218**, 804–806. (doi:10.1126/science.7134976)
- 46 Watanabe, S. 2010 Pigeons can discriminate ‘good’ and ‘bad’ paintings by children. *Anim. Cogn.* **13**, 75–85. (doi:10.1007/s10071-009-0246-8)
- 47 Watanabe, S. 2011 Discrimination of painting style and quality: pigeons use different strategies for different tasks. *Anim. Cogn.* **14**, 797–808. (doi:10.1007/s10071-011-0412-7)
- 48 Hochmann, J., Azadpour, M. & Mehler, J. 2008 Do humans really learn AⁿBⁿ artificial grammars from exemplars? *Cogn. Sci.* **32**, 1021–1036. (doi:10.1080/03640210801897849)
- 49 Perruchet, P. & Pacteau, C. 1990 Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *J. Exp. Psychol.* **119**, 264–275. (doi:10.1037/0096-3445.119.3.264)
- 50 De Vries, M. H., Monaghan, P., Knecht, S. & Zwitserlood, P. 2008 Syntactic structure and artificial grammar learning: the learnability of embedded hierarchical structures. *Cognition* **107**, 763–774. (doi:10.1016/j.cognition.2007.09.002)
- 51 Corballis, M. C. 2007 Recursion, language, and starlings. *Cogn. Sci.* **31**, 697–704. (doi:10.1080/15326900701399947)
- 52 Gerken, L. 2006 Decisions, decisions: infant language learning when multiple generalisations are possible. *Cognition* **98**, B67–B74. (doi:10.1016/j.cognition.2005.03.003)
- 53 Gerken, L. 2010 Infants use rational decision criteria for choosing among models of their input. *Cognition* **115**, 362–366. (doi:10.1016/j.cognition.2010.01.006)
- 54 Zimmerer, V. C., Cowell, P. E. & Varley, R. A. 2010 Individual behavior in learning of an artificial grammar. *Mem. Cogn.* **39**, 491–501. (doi:10.3758/s13421-010-0039-y)
- 55 Saffran, J. R., Reeck, K., Niebuhr, A. & Wilson, D. 2005 Changing the tune: the structure of the input affects infants’ use of absolute and relative pitch. *Dev. Sci.* **8**, 1–7. (doi:10.1111/j.1467-7687.2005.00387.x)
- 56 Koehler, O. 1951 The ability of birds to ‘count’. *Bull. Anim. Behav.* **9**, 41–45
- 57 Huber, L., Aust, U., Michelbach, G., Olzant, S., Loidolt, M. & Nowotny, R. 1999 Limits of symmetry conceptualization in pigeons. *Q. J. Exp. Psychol.* **52**, 351–379. (doi:10.1080/027249999393040)
- 58 Swaddle, J. P. & Ruff, A. D. 2004 Starlings have difficulty in detecting dot symmetry: implications for studying

- fluctuating asymmetry. *Behavior* **141**, 29–40. (doi:10.1163/156853904772746583)
- 59 Knowlton, B. J. & Squire, L. R. 1996 Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *J. Exp. Psychol. Learn. Mem. Cogn.* **22**, 169–181. (doi:10.1037/0278-7393.22.1.169)
- 60 Maurer, D., Le Grand, R. & Mondloch, C. J. 2002 The many faces of configural processing. *Trends Cogn. Sci.* **6**, 255–260. (doi:10.1016/S1364-6613(02)01903-4)
- 61 Jerison, J. H. 1973 *Evolution of the brain and intelligence*. New York, NY: Academic Press.
- 62 Abe, K. & Watanabe, D. 2011 Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* **14**, 1067–1074. (doi:10.1038/nn.2869)
- 63 Beckers, G. J. L., Bolhuis, J. J., Okanoya, K. & Berwick, R. C. 2012 Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* **23**, 139–145. (doi:10.1097/WNR.0b013e32834f1765)