The role of Gestalt principles in the acquisition of non-adjacent dependencies in linguistic and non-linguistic sequences

Jennifer A. Sturm (j.sturm@northumbria.ac.uk)
Kenny Smith (kenny.smith@northumbria.ac.uk)
Cognition and Communication Research Centre, Division of Psychology, Northumbria University,
Northumberland Building, Northumberland Road, Newcastle Upon Tyne, NE1 8ST, UK

Abstract

Recent evidence from artificial language learning (ALL) experiments suggests that the underlying statistical structure of the input may serve as a cue in language learning and that a similar mechanism is involved in sequential learning of non-linguistic stimuli (e.g. Christiansen, Conway & Onnis, 2007). Other experimental work suggests that Gestalt principle of similarity plays a role in the acquisition of non-adjacent dependencies (e.g. Newport & Aslin, 2004). We present experimental evidence which is inconsistent with the strongest interpretation of these previous experimental results. Adult participants in our ALL experiment learnt non-adjacent dependencies without the assistance of Gestalt principles whereas participants in equivalent non-linguistic conditions did not. This suggests, at a minimum, that any domain-general learning capacity employed in language acquisition must be provided with domain-specific expectations about the relevant units of analysis.

Keywords: artificial grammar learning, non-adjacent dependencies, Gestalt principle of similarity

Introduction

The question of fundamental importance in the area of language learning concerns the relative ease with which humans acquire their native language. The grammatical structure of natural languages is immensely complex and yet humans are able to learn the underlying grammatical structure of their native language from an incomplete and imperfect input (Chomsky, 1965). Recent interest has focused on whether the processes involved in language acquisition are part of a more general learning mechanism also responsible for non-linguistic sequence learning (Gómez & Gerken, 2000; Saffran, Johnson, Aslin & Newport, 1999, Christiansen & Chater, 2008).

Non-adjacent dependencies pose a challenge for language learners, since they are required to identify a relationship between units separated by a (potentially arbitrary) number of intervening units. Two examples of non-adjacent dependencies frequently occurring in English are given below: subject-verb agreement (1) and wh-dependencies (2). In both cases learners have to track the dependencies between underlined elements across an arbitrary number of other words.

(1) *The cheese in the fridge is mouldy.*
(2) *What did John do_?*

Although non-adjacent dependencies are ubiquitous in natural languages, the ability of humans to learn them in ALL experiments seems to be subject to specific constraints. As discussed below, humans have been shown to detect non-adjacent regularities in linguistic and non-linguistic stimuli when Gestalt principles assist in making the regularities more salient.

Non-adjacent dependencies in ALL experiments

The Gestalt theory suggests that stimuli are grouped together according to organizing principles. One of these principles is the law of similarity, meaning that materials that are perceived to be similar are more readily grouped together no matter what their spatial or temporal relationship might be (Wertheimer, 1938).

Existing ALL studies of non-adjacent dependency learning suggest that people rely on Gestalt principles as a cue as to which elements form the dependency. Newport and Aslin (2004) investigated the constraints on learning non-adjacent dependencies. In their experiments, adult participants were successful at learning regularities between segments only, more specifically dependencies between consonants when the intervening element was a vowel, and dependencies between vowels, skipping consonants. Newport and Aslin suggest Gestalt principles as a possible explanation for why humans would be able to compute these non-adjacent dependencies between segments, but perform poorly at detecting regularities between other non-adjacent units, i.e. syllables. In line with the Gestalt principle of similarity, people are more readily capable of detecting regularities between segments since all vowels share common properties, as do all consonants.

Earlier work by Gómez (2002) also suggests a role for Gestalt principles in the learning of non-adjacent dependencies. Gómez demonstrates that both human adults and 18 month-old human infants are able to learn non-adjacent dependencies in artificial languages when the transitional probabilities of adjacent words in the input are not reliable enough to extract rules. She familiarized participants with sequences consisting of three nonsense words, where there was a dependency between the first and the final word, and the middle element varied freely (see Table 1, which gives the grammar used for adult participants; infants were trained on a slightly simpler

grammar). Elements a – f were represented by the monosyllabic CVC words (*pel, vot, dak, rud, jic, tood),* whereas the X elements were bisyllabic (e.g. *hiftam, benez).*

Table 1: Gómez's grammar.

Language 1	Language 2
$S_1 \rightarrow$ aXd	$S_1 \rightarrow$ aXe
$S_2 \rightarrow$ bXe	$S_2 \rightarrow$ bXf
$S_3 \rightarrow$ cXf	$S_3 \rightarrow$ cXd

The familiarization phase was followed by a testing phase, in which participants were asked to distinguish between grammatical and ungrammatical strings: ungrammatical strings for participants trained on L1 were the grammatical strings from L2 and vice versa. Gómez found that the more variable the middle element (i.e. the bigger the pool from which the X element is drawn), the more likely people were to learn the non-adjacent dependencies. Gómez suggests that learners seek invariant structure in the input: high variability of the intervening unit makes the transitional probabilities between adjacent words so unreliable that participants reject the idea of adjacent dependencies and focus their attention on regularities between non-adjacent elements.

Gómez facilitates the detection of the non-adjacent dependencies by exploiting the Gestalt principle of similarity. In her AL, the words involved in the dependencies were monosyllabic, whereas the words belonging to category X were bisyllabic. Gómez therefore includes a cue which participants could use to identify the dependencies, or at least a cue which highlights the elements over which the dependency operates.

Non-adjacent dependencies and non-linguistic stimuli

The Gestalt principle of similarity seems to also facilitate detection of non-adjacent dependencies in the non-linguistic domain, as has been shown by Creel, Newport and Aslin (2004). In their series of experiments, participants were able to acquire non-adjacent regularities between aurally presented tone sequences as long as the elements forming the dependencies were similar in pitch or timbre. Thus, tones seem to be more readily grouped together due to their featural similarity, even if they are not temporally adjacent. These results indicate that, with simple patterns underlying simple stimuli, a domain-general learning mechanism might be at work.

Other work looking at equivalences between sequential learning of linguistic and non-linguistic sequences also suggests a potential role for Gestalt principles. To take one example: Kirkham, Slemmer and Johnson (2002) demonstrate that probabilistic cues (element-to-element transitional probabilities: Saffran, Aslin & Newport, 1996) used for sequence segmentation of linguistic stimuli can also be applied to segment non-linguistic sequences. In

their original experiment, Saffran et al. showed that infants are aware of the syllable-to-syllable transitional probabilities of a briefly-presented auditory stimulus, and can subsequently differentiate between sequences involving high and low-probability transitions. Kirkham et al. (2002) demonstrate the same result pertains in the visual domain, when infants are trained and tested on sequences of colored geometrical shapes.

However, the visual stimuli in this experiment do not correspond directly to the linguistic stimuli used by Saffran et al. (1996). Saffran et al.'s linguistic stimuli involve combinatorial reuse of consonants and vowels (e.g., *golabu* and *bidaku* share the plosive *b* and two vowels). The discrete shapes used by Kirkham et al. are non-combinatorial, in that each word in the Saffran et al. stimuli is replaced by a geometrical shape which differs in both shape and color from the other shapes (i.e. there is no re-use of shape or color across the shapes corresponding to *golabu* and *bidaku)*, and are therefore less complex than the equivalent linguistic stimuli used by Saffran et al. (1996).

Experiments in the area of sequential learning of non-adjacent dependencies carried out to date therefore suggest the following three research questions: (1) Are people capable of detecting non-adjacent dependencies between linguistic elements without assistance from the Gestalt principle of similarity?; (2) Are the mechanisms involved in identifying non-adjacent dependencies language-specific or do they form part of a more general inventory of learning tools?; (3) In line with our first research question, is the acquisition of non-adjacent dependencies between complex non-linguistic patterns possible when the stimuli highlight the relevant units over which the dependencies should operate?

Experiment

We tested human adults' ability in detecting non-adjacent dependencies in two different domains: the linguistic domain was realised using an AL (*language* condition), and the non-linguistic domain using two sets of black and white matrix patterns (the *componential* and *holistic* conditions, described below). Based on the experiments discussed above, our prediction was that all three conditions should elicit the same results: although not making use of the Gestalt principle would make learning the dependency more difficult, the literature in the field of AL strongly suggests that performance in the linguistic and non-linguistic domain should not differ.

Method

Participants. Ninety-eight adults were recruited from the undergraduate population at Northumbria University and from our research centre's pool of regular experimental participants. They participated for either course credit or £ 4.50. All participants were English native speakers. Two participants were excluded from analysis as they took part in two of three conditions, the remaining participants were

evenly distributed across conditions (32 participants per condition).

Materials. The experiment consisted of three conditions differing in instructions and the materials. The grammar used for all three conditions was based on the grammar from Gómez (2002) for the largest set size of 24 (see Table 2).

Table 2: The underlying grammar (top) and lexical items used in the language condition (below).

$S_1 \rightarrow aXd$	
$S_2 \rightarrow bXe$	
$S_3 \rightarrow cXf$	
L1	**L2**
a → lum d → fip	a → nis d → huk
b → zel e → pof	b → jad e → zin
c → vok f → gam	c → fet f → gos
X → {*fet, fub, fum, gos, huk, hup, jad, jeg, lek, lep, lig, lof, lud, nis, nug, nup, pif, pir, taf, vam, vek, zec, zin, zog}*	X → { *fip, fub, fum, gam, hup, jeg, lek, lep, lig, lof, lud, lum, nug, nup, pif, pir, pof, taf, vam, vek, vok, zec, zel, zog}*

Linguistic stimuli: In the language condition, categories were instantiated as words (see Table 2). Unlike in Gómez's AL, both the words involved in the dependency and the intervening X elements were monosyllabic CVC words – this eliminates the potential for the Gestalt principle to highlight the elements over which the dependency should operate. The two languages differed in their assignment of words to categories, to control for arbitrary preferences for specific elements.

Componential non-linguistic stimuli: For the first of the non-linguistic conditions, the AL used in the language condition was converted into complex black and white matrix patterns, where there was a direct correspondence between orthographic characters in the linguistic stimuli and sub-components of the matrix patterns: each letter was directly translated into a pattern of black and white cells, rendering a complex matrix for each word whose internal structure corresponds to the internal structure of the words used in language condition. So, for instance, every "l" from each of the words in the language condition corresponds to a certain pattern within a grid, as does ever "u" and every "m". An example for these three letters and their corresponding patterns for condition 2 are shown in Fig. 1.


Figure 1: Translation of stimuli in the componential condition.

Holistic non-linguistic stimuli: The stimuli used for the third condition (holistic condition) were also black and white matrix patterns as used in the componential condition, however, the patterns in this case have less complex internal structures, and were generated to appear more like a single unit rather than a pattern composed of three individual units. Every occurrence of each word from the language condition was mapped onto a distinct pattern. Fig.2 shows an example of one string in all three conditions.


Figure 2: Example stimuli.

Stimuli in all three conditions were presented visually, on a white computer screen. The three elements that formed one sequence were displayed simultaneously for 2500 ms, and each presentation of a string was separated by 1000 ms pauses (i.e. blank screen) from the next one. While simultaneous visual presentation is a significant departure from Gómez's sequential auditory presentation, Saffran (2002) suggests that simultaneous presentation facilitates the detection of regularities within the input. The experiment was designed using the software package Slide Generator .[1]

Procedure. The experiment consisted of an initial training phase, in which participants were exposed to either L1 or L2. In both languages, each of the 24 X elements appeared in each of the three dependencies three times in random order, rendering a total of 216 (3 dependencies x 24 X elements x 3 repetitions) sequences. This phase lasted approximately 20 minutes. Each participant was merely asked to pay careful attention during their exposure to a large number of sequences consisting of either three words (in the language condition) or three patterns (in the non-linguistic conditions), as they were going to be tested on these sequences later on.

Before the testing phase started, participants were informed that the sequences they had been exposed to during training had followed specific rules, and that for each sequence that appeared on the screen in the testing phase, they had to decide whether or not it followed the same rules as the sequences from the training phase. The participants indicated their response by key press with their dominant index finger – "yes" if they thought the sequence followed the same rule, "no" if it didn`t. The V and B keys on the

---

keyboard served as the "yes" and "no" keys and were therefore marked with either "Y" or "N", counterbalanced across participants. The testing phase took between 20 and 25 minutes.

Unlike in Gómez (2002), participants were tested on sequences involving both familiar and novel X elements. Half of the X elements participants were trained on were replaced by novel tokens. Participants trained on grammar L1 were split into two sub-groups on test, L1a and L1b, with L1a and L1b differing in which familiar X elements were replaced with novel X elements. Participants trained on L2 were similarly sub-divided on test. The illegal endpoints for the grammatical violations varied between the sub-conditions (see Table 3). Grammatical strings followed the grammar shown in Table 2.

Table 3: Violations for Testing Phase.

Version a	Version b
aXf	aXe
bXd	bXf
cXe	cXd

Results

The percentage of test sequences which were endorsed was contrasted across conditions, using an analysis of variance with stimuli type (language, componential and holistic), test condition (version a and version b) and language (L1 and L2) as between-subject factors, and familiarity with X (familiar and unfamiliar) and grammaticality as within-subjects factors. The ANOVA showed a main effect of grammaticality, $F(1, 90) = 15.29$, $p < .001$ and, more relevantly, a grammaticality x stimuli type interaction, $F(2, 90) = 4.99$, $p = .009$ (see Fig. 3).
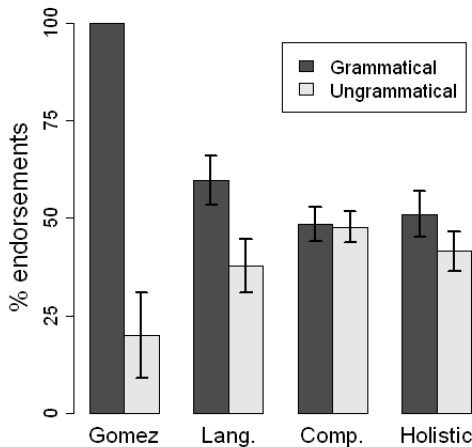

Figure 3: Our results for each condition contrasted directly with Gómez`s findings.

This interaction was further investigated by running paired-samples t-tests for each condition, comparing each

participants' endorsements of grammatical strings with their endorsements of ungrammatical strings. Only participants in the language condition reliably discriminated between grammatical and ungrammatical strings ($t(31) = 3.605$, $p = .001$). There was no significant result in either of the non-linguistic conditions (componential: $t(31) = 0.395$, $p = 0.695$; holistic: $t(31) = 1.889$, $p = 0.068$), although, the holistic condition was closer to significance than the componential condition. The omnibus ANOVA also resulted in a main effect of familiarity of the X items, $F(1, 90) = 87.24$, $p < .001$ and a familiarity x stimuli type interaction, $F(2, 90) = 8.16$, $p = .001$. The main effect reflects the fact that participants were more willing to accept sequences containing X elements they had encountered during the training phase (see Fig. 4). The interaction indicates that, regardless of grammaticality, participants in the holistic condition were particularly unlikely to accept unfamiliar X elements (mean difference between endorsements for familiar and novel X sequences = 20.313 in the holistic condition, compared to 11.157 in the language condition, 7.031 in the componential condition). This shows that in the holistic condition participants were relying on memorising sequences more than in the language condition, where general rules were extracted.
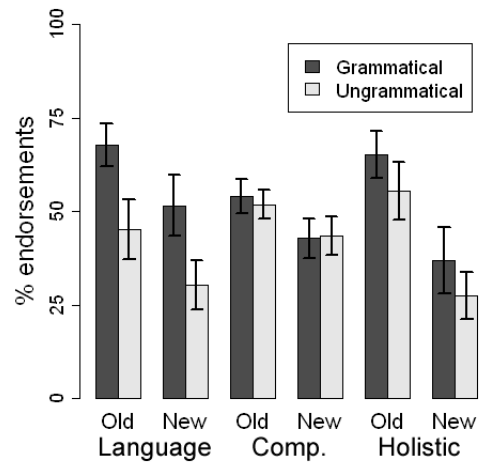

Figure 4: Endorsements with grammatical and ungrammatical strings containing old (familiar) and new (unfamiliar) X elements for each condition.

The ANOVA also resulted in a significant between-subjects effect for testing condition (a versus b), $F(1, 84) = 4.477$, $p = .037$, with participants in test condition a providing more endorsements (endorsing 49.6% of sequences, compared to 45.8% for participants in test condition b). However, while the interpretation of this finding is not clear, we do not consider it to be important, for two reasons. Firstly, the interaction between this factor and grammaticality was not significant ($F = 2.333$, $p = .130$), suggesting that this difference does not reflect a difference in ability to discriminate grammatical and ungrammatical sequences. Secondly, the test condition x

stimuli type interaction was also not significant (F = 2.060, p = .134). Given that the patterns used in the holistic condition were arbitrarily assigned to words from the language condition, if the main effect for test condition reflected some facilitatory selection of words in the language and componential conditions (e.g. maybe there was some unforeseen similarity between novel X items and retained familiar X items) we would not expect to see this effect in the holistic condition. There were no other significant main effects or interactions.

As can be seen in Figs. 3 - 4, in the language condition, participants accepted grammatical strings approximately 20% more often than ungrammatical strings, regardless of whether the X element was familiar or novel. The general pattern of results for the holistic condition resembles the language condition: the discrepancy between endorsements with grammatical and ungrammatical strings remains constant in spite of familiarity or unfamiliarity with X, in this case approximately 10%. The pattern of endorsements in the componential condition is rather different, again suggesting that participants in this condition did not acquire the grammar and were thus not able to apply the rules to the stimuli in the testing phase.

Overall, these data suggest that the underlying grammar was only recognised in the language condition. The inclusion of unfamiliar X elements in the testing phase played an important role in judgements of sequences in all three conditions, with participants generally being less willing to endorse sequences involving unfamiliar X items.

How many non-adjacent dependencies were our participants able to learn? We can attempt to answer this by looking in more detail at the correct responses given for each dependency by each participant. Participants were tested on each of the three dependencies 48 times, 24 times when the dependency was observed and 24 times when it was violated. According to the binomial, 31 correct responses ("Y" for grammatical and "N" for ungrammatical sequences) out of 48 test reflects a level of performance unlikely to be achieved by chance (p = 0.0297). We therefore classified each dependency as learned by a participant if they scored 31 or above on testing on that dependency. The results of this test applied to every participant are shown in Table 5. According to this criterion, 17 of 32 participants in the language condition mastered at least one of the dependencies, whereas the majority of all participants in the two non-linguistic conditions failed to learn any of the regularities.

Table 5: Number of dependencies learnt.

	# dependencies learnt			
Condition	0	1	2	3
Language	15	8	3	6
Componential	25	7	0	0
Holistic	26	3	0	3

Discussion

The purpose of this series of studies was twofold: We further investigated the question of whether the ability of humans to compute non-adjacent dependencies in a linguistic input is part of a domain-general learning device. In doing this, we simultaneously examined the importance of the Gestalt principle in detecting non-adjacent regularities.

If being able to track these regularities were indeed a fully domain-general capacity, we would expect participants to perform equally well in the linguistic condition and in the non-linguistic conditions. However, the behavioral data collected here do not reflect this. In this respect, our results do not conform to the majority of ALL literature, in which the general consensus seems to be that the same underlying mechanisms are used for sequential learning, regardless of the domain (see Christiansen et al., 2007; Kirkham et al., 2002). Instead, the results here suggest that – at a minimum – people have modality-specific expectations.

Unlike in previous experiments (Kirkham et al., 2002; Creel et al., 2004), the non-linguistic stimuli used in our experiments reflect the complexity of the linguistic stimuli in that the individual words consist of three parts (i.e. letters), as do the matrix patterns in the componential condition. This was not the case in Kirkham et al.`s visual materials, and the non-adjacent dependencies in the auditory stimuli used by Creel et al. were also between single tones lacking the element of componentiality of words.

In our series of experiments, our participants reliably distinguished between grammatical and ungrammatical sequences in the language condition but not in the two non-linguistic conditions. A possible explanation for this finding is that people seem unable to detect the regularities underlying the sequential non-linguistic input due to the fact that they pay too much attention to the internal structure of the matrix patterns, trying to find regularities within the structures themselves. This is not the case in the language condition, where more than 50% of participants detected at least one non-adjacent dependency (see Table 5). In the language condition, participants seem less likely to get lost in detail, perhaps employing a heuristic such as "ignore the internal structure of each of the words as regularities between individual letters do not play an important role in English", indicating that people might have language-specific expectations.

Our attempt to assist participants in learning non-linguistic non-adjacent dependencies, by designing holistic visual patterns and thereby eliminating the very detailed internal structure in the componential visual patterns, resulted in performances lying somewhere between the language and the componential condition. This finding further supports the notion of modality-specific expectations: for linguistic stimuli, people seem to be aware of the kinds of regularities to ignore and ones to focus on. By giving participants less internal structure to deal with, we facilitated the shift of focus onto regularities between (rather than within) individual units. However, since our results do not show an

effect of grammaticality in the holistic condition, the patterns are not quite simple enough for people to ignore the internal structure completely. Nevertheless, our prediction is that if we translated our AL into the non-linguistic domain using the very simple shapes Kirkham et al. (2002) used, then the results would not significantly differ from our results in the language condition. Note, however, that this requires that we use non-linguistic stimuli which do not match the complexity of the linguistic stimuli, in order to compensate for the different prior expectations of learners in these two domains.

In general, our data suggest that non-linguistic visual stimuli of comparable complexity to linguistic materials render significantly different results. This indicates that the processes involved in computing non-adjacent dependencies in the linguistic and non-linguistic domain are not exactly the same. At a minimum, learners bring different prior expectations about the relevant units of analysis to learning tasks in different domains.

In terms of the Gestalt principle of similarity, our results suggest that people can indeed detect non-adjacent regularities within a linguistic input even when the relevant words do not share more common properties than they do with the intervening word. Unlike in Gómez (2002), in our language condition, participants did not have a salient length cue highlighting the units involved in the dependency. It is therefore not surprising that our results differ from Gómez`s in as much as that 15 of 32 participants in our language condition were non-learners (see also Fig. 3).

There are two factors that might explain the discrepancy between Gómez`s and our results. Firstly, the Gestalt principle of similarity may have made our AL significantly more difficult to learn, as explained above. Secondly, the modality of the input may play a role. Gómez presented the stimuli aurally, whereas all our materials were presented visually. In doing this, Gómez may be tapping into particularly strong (possibly learned) expectations about the relevant units of analysis in auditory sequences, and even the possibility of non-adjacent dependencies. To what extent the choice of modality plays a role in sequential learning of non-adjacent dependencies is worth investigating in further experiments.

Conclusion

Our results show that not only do modality-specific expectations affect performance in processing non-adjacent dependencies, but also that non-adjacent dependencies in the linguistic domain can be acquired without facilitating Gestalt cues. While humans are capable of learning non-adjacent dependencies in a linguistic input, they are not able to acquire the same grammar when the stimuli are in the non-linguistic domain where stimuli are closely matched in terms of internal complexity. This suggests that specifically linguistic learning mechanisms (or specifically linguistic expectations feeding into a domain-general mechanism) assist in detecting these regularities. To date, the human ability to learn non-adjacent dependencies has been assumed to be constrained by the Gestalt principle of similarity. Our results indicate that while Gestalt cues may facilitate the acquisition of these dependencies, they are not necessary. The majority of our participants were able to identify non-adjacent regularities in the absence of such cues.

Acknowledgements

References

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

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

Christiansen, M.H., Conway, C. M., & Onnis, L. (2007) Neural responses to structural incongruencies in language and statistical learning point to a similar underlying mechanism. In D.S. McNamara & J.G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp.173-178). Austin, TX: Cognitive Science Society.

Creel, S.C., Newport, E.L., & Aslin, R.N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 1119–1130.

Fiser, J., & Aslin, R.N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 458-467.

Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.

Gómez, R., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178-186.

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.

Newport, E.L., & Aslin, R.N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27 – 52.

Saffran, J.R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172-196.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. Ellis (Ed. And Trans.), *A source book of Gestalt psychology*. London: Routledge & Kegan Paul.