



This article is part of the topic “Learning Grammatical Structures: Developmental, Cross-species and Computational Approaches,” Carel ten Cate, Clara Levelt, Judit Gervain, Chris Petkov and Willem Zuidema (Topic Editors). For a full listing of topic papers, see [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1756-8765/earlyview](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview)

Non-adjacent Dependency Learning in Humans and Other Animals

Benjamin Wilson,^a Michelle Spierings,^b Andrea Ravignani,^{c,d}
Jutta L. Mueller,^e Toben H. Mintz,^f Frank Wijnen,^g Anne van der Kant,^h
Kenny Smith,ⁱ Arnaud Rey^j

^a*Institute of Neuroscience, Newcastle University*

^b*Department of Cognitive Biology, University of Vienna*

^c*Research Department, Sealcentre Pieterburen*

^d*Artificial Intelligence Lab, Vrije Universiteit Brussel*

^e*Institute of Cognitive Science, University of Osnabrueck*

^f*Departments of Psychology and Linguistics, University of Southern California*

^g*Utrecht Institute of Linguistics OTS, Utrecht University*

^h*Department of Linguistics, University of Potsdam*

ⁱ*Centre for Language Evolution, University of Edinburgh*

^j*CNRS, Aix-Marseille University*

Received 26 February 2018; received in revised form 22 May 2018; accepted 30 May 2018

Abstract

Learning and processing natural language requires the ability to track syntactic relationships between words and phrases in a sentence, which are often separated by intervening material. These nonadjacent dependencies can be studied using artificial grammar learning paradigms and structured sequence processing tasks. These approaches have been used to demonstrate that human adults, infants and some nonhuman animals are able to detect and learn dependencies between nonadjacent elements within a sequence. However, learning nonadjacent dependencies appears to be more cognitively demanding than detecting dependencies between adjacent elements, and only

Correspondence should be sent to Benjamin Wilson, Institute of Neuroscience, Newcastle University Medical School, Newcastle Upon Tyne, NE2 4HH, UK. E-mail: benjamin.wilson@newcastle.ac.uk

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

occurs in certain circumstances. In this review, we discuss different types of nonadjacent dependencies in language and in artificial grammar learning experiments, and how these differences might impact learning. We summarize different types of perceptual cues that facilitate learning, by highlighting the relationship between dependent elements bringing them closer together either physically, attentionally, or perceptually. Finally, we review artificial grammar learning experiments in human adults, infants, and nonhuman animals, and discuss how similarities and differences observed across these groups can provide insights into how language is learned across development and how these language-related abilities might have evolved.

Keywords: Non-adjacent dependency; Artificial grammar; Structured sequence processing; Human; Infant; Nonhuman animal; Primate

1. Introduction

A central feature of syntactic processing is the ability to track structural relationships between words and phrases in a sentence. However, language has a hierarchical structure; hence, syntactic relationships exist not only between *adjacent words* but also across longer distances, requiring the joint processing of words separated by intervening material. Tracking such *nonadjacent dependencies* is more cognitively complex than processing the relationships between adjacent words, minimally placing additional demands on working memory. Moreover, the ability to detect nonadjacent dependencies is a crucial prerequisite for learning and processing some more complex syntactic relationships (like center embedded structures). Understanding how we learn to detect and process nonadjacent dependencies, and the ontogenetic and evolutionary origins of these abilities, represents an important challenge in understanding key aspects of language processing, acquisition, and evolution.

A nonadjacent dependency is a relationship between two temporally or spatially separated elements, which cannot be explained simply by the occurrence of multiple pairs of adjacent relationships. For example, in English, nonadjacent dependencies can be observed in tense agreement (e.g., “Is [talk]ing,” “Has [talk]ed”) or subject-verb agreement (e.g., “the dog [down the street] barks” vs. “the dogs [down the street] bark”). In both of these cases the dependent elements can be separated by intervening material, and one must track the nonadjacent dependency, holding the first element in memory until the appropriate point later in the sequence.

One productive way to study how dependencies are learned and processed is to use artificial grammar learning paradigms and structured sequence processing tasks. These experiments typically test participants’ abilities to learn relationships between specific elements in sequences of auditory or visual stimuli (Reber, 1967). Such tasks do not rely on existing semantic or syntactic knowledge and are therefore an important tool with which to isolate and study the cognitive and neurobiological systems that support how specific aspects of syntax may be learned and processed (Petersson, Folia, & Hagoort, 2012). Furthermore, these tasks seem to tap into language-relevant capacities,

as evidenced by correlations between nonadjacent dependency learning and natural language processing in adult humans (Misyak, Christiansen, & Tomblin, 2010), and by the fact that nonadjacent dependency learning is impaired in individuals with specific language impairment (Hsu, Tomblin, & Christiansen, 2014) or familial risk of dyslexia (Kerkhoff, De Bree, De Klerk, & Wijnen, 2013). Finally, as these tasks do not require language, they can be used to test preverbal infants, to inform us about language acquisition (e.g., Gómez, 2002; Saffran, Aslin, & Newport, 1996), and non-human animals, which can provide insights into language origins and evolution (e.g., Fitch & Hauser, 2004; Newport, Hauser, Spaepen, & Aslin, 2004).

In a seminal study of nonadjacent dependency learning, Gómez (2002) presented adult participants and 18-month-old infants with sequences of nonsense words (e.g., “*pel wadim rud*,” “*vot kicey jic*”) containing a nonadjacent dependency in which the first word predicted the final word (i.e., “*pel*” predicts “*rud*”; “*vot*” predicts “*jic*” regardless of the identity of the middle word in the sequence). Participants were then tested with sequences that were either consistent or inconsistent with this dependency (e.g., “*pel wadim rud*” vs. “*pel wadim jic*”). Gómez found that in some conditions (see below) adults and infants are able to learn these nonadjacent dependencies, discriminating between sequences which conform to or violate the dependency. This result has been confirmed in further experiments involving adults (e.g., Frost & Monaghan, 2016; Gómez, 2002; Pena, Bonatti, Nespor, & Mehler, 2002; van den Bos, Christiansen, & Misyak, 2012; Vuong, Meyer, & Christiansen, 2016) and infants (e.g., Gómez, 2002; Gómez & Maye, 2005) as well as nonhuman animals (e.g., Milne et al., 2016; Newport et al., 2004; Ravignani, Sonnweber, Stobbe, & Fitch, 2013; Sonnweber, Ravignani, & Fitch, 2015; Versace, Rogge, Shelton-May, & Ravignani, 2017), although successful learning only occurs under certain conditions (see Section 3, below).

In this article, we first characterize different types of nonadjacent dependencies, and how they relate to nonadjacent syntactic relationships in natural languages. We summarize the evidence for nonadjacent dependency learning in human adults and infants, and nonhuman animals, review the conditions under which learning does and does not occur, and discuss the insights these studies provide into the development and evolution of these abilities.

2. Characterizing different types of nonadjacent dependency

The simplest form of structural dependency that might appear within a sentence is a syntactic relationship between two adjacent words. In artificial grammar learning tasks, this type of adjacent dependency might be present between two sequentially presented elements “A” and “B” (see Fig. 1i). These adjacent dependencies are present in most artificial grammar learning studies and are rapidly learned by human adults and infants, and nonhuman animals (e.g., Chen & ten Cate, 2015; Fitch & Hauser, 2004; Gebhart, Newport, & Aslin, 2009; Pacton, Sobaco, & Perruchet, 2015; Reber, 1967; Saffran et al., 1996; Wilson, Smith, & Petkov, 2015). By contrast, a nonadjacent dependency represents a relationship between items that are separated by one or more intervening elements in a

sequence (Fig. 1ii and iii, e.g., Gómez, 2002; Newport et al., 2004). Both adjacent and nonadjacent dependencies might require participants to learn the relationship between the same two elements (“A” and “B” in Fig. 1), yet the cognitive and memory demands might vary in relation to the distance over which this dependency must be processed. Such nonadjacent relationships might appear within a longer sequence (Fig. 1iv), removing the opportunity to rely on the edges of the sequence to identify key dependencies (e.g., Endress, Carden, Versace, & Hauser, 2010; Endress, Nespors, & Mehler, 2009; Grama, Wijnen, & Kerkhoff, 2013). Finally, many artificial grammar learning paradigms are not designed specifically to assess the learning of nonadjacent dependencies, but nevertheless require this ability as a prerequisite to process more complex dependencies. For example, many studies have used a grammar of the form A^nB^n to assess different forms of hierarchical sequence processing (e.g., center-embedding, crossed dependencies, which are present in some natural languages [Bach, Brown, & Marslen-Wilson, 1986], Fig. 1v, vi), which require the participant to process multiple adjacent and nonadjacent dependencies between “A” and “B” elements (De Vries, Christiansen, & Petersson, 2011; De Vries, Petersson, Geukes, Zwitserlood, & Christiansen, 2012). While these more complex tasks bring additional cognitive demands, all of these tasks require the ability to identify and process nonadjacent dependencies, which represents an important prerequisite to aspects of syntax processing.

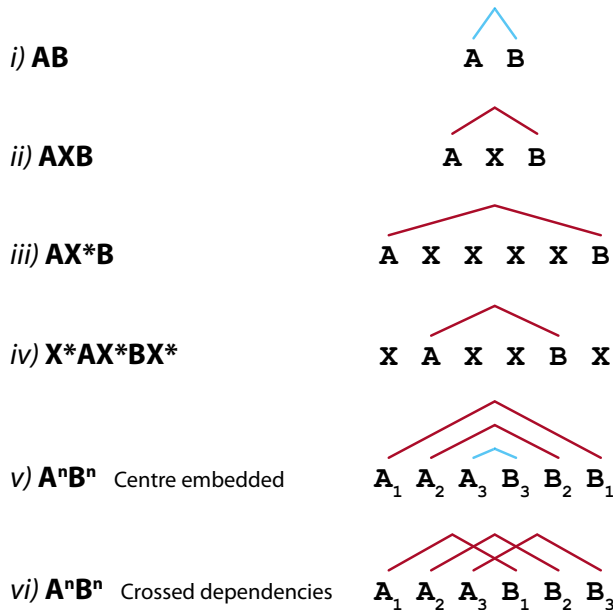


Fig. 1. Adjacent and nonadjacent dependencies in several commonly used artificial grammar structures. Adjacent dependencies are shown in blue and nonadjacent dependencies are shown in red. More complex grammars (e.g., A^nB^n) go beyond the requirement to learn a single nonadjacent dependency at a time and require several dependencies to be processed simultaneously.

Sequence learning experiments have also manipulated the nature of the relationships between the dependent elements. A simple type of dependency is the relationship between two identical elements, and it is known as an identity relation (see Fig. 2i). This type of identity relation is sometimes observed in natural language. For example, agreement systems in Bantu languages often involve the appearance of an identical prefix on words which agree, as in Swahili, for example, the “*ki*” prefix in “*Ki-kapu ki-dogo ki-me-fika*” (“the little basket arrived”). It has been suggested that this transparent means of signaling agreement in these languages allows them to sustain unusually complex agreement systems arising from a large number of noun classes (Demuth & Weschler, 2012). In an artificial grammar learning task, a nonadjacent dependency of this type might take the form “AXA,” where a dependency exists between two identical sequence elements. Learning and processing this dependency requires only that the initial element be held in memory, and that incoming stimuli be compared to this memory representation.

Bantu languages like Swahili notwithstanding, most languages rarely require us to detect repetitions of the same element within a sentence (as in Fig. 2i), but rather to track dependencies between different words or part-words that share little phonological resemblance, based on learned relationships. As in the previous example of tense agreement,

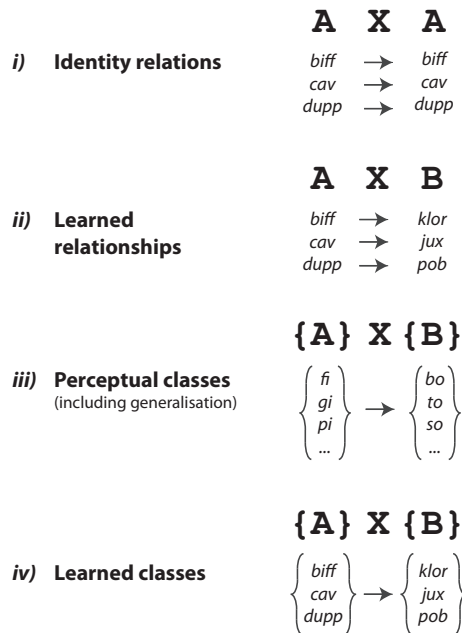


Fig. 2. Different types of stimulus classes used in nonadjacent dependency learning tasks. Several varieties of nonadjacent dependencies can be assessed in artificial grammar learning studies. These include (i) identity relations between specific elements (as can be seen in certain Bantu languages); (ii) learned relationships between specific elements (as in English tense agreement); (iii) relationships between perceptual classes; (iv) relationships between learned classes (similar to dependencies between syntactic word categories; for example, nouns and verbs). See Endress and Bonatti (2007).

one must learn the relationship between “*is*” and the suffix “*_ing*” based on their co-occurrence, before one is able to process this dependency or detect ungrammatical sentences (e.g., Friederici, Mueller, & Oberecker, 2011). This has been investigated using artificial grammars of the form “AXB” (Fig. 2ii). In such studies the first element predicts a different final element, and the dependent elements are related only by an arbitrary pairwise association, which must first be learned by the participants (e.g., Gómez, 2002).

Apart from relations between specific linguistic elements, natural languages feature dependencies between syntactic classes or categories of words; for example, subject-verb agreement in English requires subjects and verbs to agree for number: singular subjects must appear with singular verbs, regardless of the specific identity of the nouns and verbs involved. In these cases, it is necessary to both learn which category a specific word belongs to and to also learn the relationships between these categories. This can be assessed using an artificial grammar of the form “{A}X{B},” where {A} and {B} represent two sets of stimuli and therefore any stimulus from set {A} predicts any {B} category stimulus (Fig. 2iii,iv). To limit the requirement that participants must learn the categories associated with many different stimuli, some studies have used sets of stimuli in which perceptual cues denote category membership (see Fig. 2iii). For example, all {A} stimuli might be nonsense syllables containing the vowel “*i*,” while {B} stimuli contain “*o*” vowels, thus adding a clear perceptual cue to the stimulus categories. A participant would be required to first recognize that there are two (perceptually different) categories of stimuli, and then to recognize and learn the nonadjacent dependency between them (in this case, “*_i*” syllables predict “*_o*” syllables; see Fig. 2iii). This approach has been used to assess the learning of center embedded structures, which include nonadjacent dependencies (see Fig. 1v) in humans and nonhuman animals (Bahmann, Schubotz, & Friederici, 2008; Fitch & Hauser, 2004; Friederici, Bahmann, Heim, Schubotz, & Anwander, 2006). One advantage of this approach is that once participants have learned these relationships, they may be able to generalize to novel stimuli as long as these can also be categorized using the perceptual cue.

Clear operationalized descriptions of specific types of nonadjacent dependencies (Fig. 2) may help us to understand how different types of dependencies are learned and processed across different populations and species. This clarity may be important in understanding the developmental and evolutionary origins of abilities that are critical to language.

3. Nonadjacent dependency learning in adults, infants, and nonhuman animals

In the previous section, we summarized several different types of nonadjacent dependencies found in natural language and how they have been studied using artificial grammar learning tasks. Next, we review how human adults and infants and nonhuman animals have been tested using these paradigms to better understand how different types of dependencies (Figs. 1 and 2) are learned and how these might vary across development and evolution.

3.1. Human adults and infants

Infants as young as 7 months old readily notice dependencies between identical nonadjacent items in a sequence (e.g., “*ga po ga*” vs. “*ga po bi*,” represented as “AXA” in Fig. 2i) (Gerken, 2006; Gervain & Werker, 2013; Marcus, Vijayan, Rao, & Vishton, 1999). As we discussed above, these identity relations might be relatively easy to recognize because they do not require the participant to learn an arbitrary relationship between two different stimuli (as in Fig. 2ii,iii).

Many studies have also assessed the learning of dependencies between arbitrarily related stimuli, which might be relevant to a wider range of nonadjacent dependencies seen in natural languages. Arbitrary item-based dependencies between adjacent elements appear to be easily learned from shortly after birth (e.g., Aslin, Saffran, & Newport, 1998; Kudo, Nonaka, Mizuno, Mizuno, & Okanoya, 2011; Perruchet & Pacton, 2006; Saffran, Johnson, Aslin, & Newport, 1999; Saffran et al., 1996; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). Furthermore, several studies have found that both adults and infants as young as 3 months old show learning of nonadjacent dependencies of the form “AXB” (e.g., Citron, Oberecker, Friederici, & Mueller, 2011; Frost & Monaghan, 2016; Gómez, 2002; Gómez & Maye, 2005; Marchetto & Bonatti, 2013; Mueller, Friederici, & Mannel, 2012; Mueller, Oberecker, & Friederici, 2009; Pena et al., 2002; van den Bos et al., 2012; Vuong et al., 2016). However, as we discuss below, subsequent work has revealed important limits on the human capacity for learning nonadjacent dependencies and identified a number of additional cues that might emphasize these relationships and aid learning (see below and Fig. 3).

3.1.1. Variability and the detection of predictable dependencies

Gómez (2002) showed that both adults and 18-month-old infants were able to detect a nonadjacent dependency between the first and last items of a three-element sequence “AXB,” where the dependency involved an arbitrary association between items (e.g., initial “*pel*” predicted final “*rud*”). However, in both groups learning only occurred when a large number of different nonsense words were presented in the second position in the sequence, making the adjacent dependencies between the first and second, and second and third elements highly unpredictable (Fig. 3). When a smaller number of “X” elements was used, participants appeared to try to learn the (uninformative) adjacent relationships and failed to learn the nonadjacent dependency (e.g., Gómez, 2002; Gómez & Maye, 2005). This suggests that, while both adults and infants are able to implicitly learn nonadjacent dependencies, in some cases they may fail to do so if more salient adjacent cues are present. However, in some cases both adjacent and nonadjacent dependencies may be learned simultaneously (Romberg & Saffran, 2013b; Wang & Mintz, 2018), at least in some participants (see Milne, Petkov, & Wilson, 2017; Wilson et al., 2015), suggesting that the processes involved in learning these two types of relationships are not inherently antagonistic.

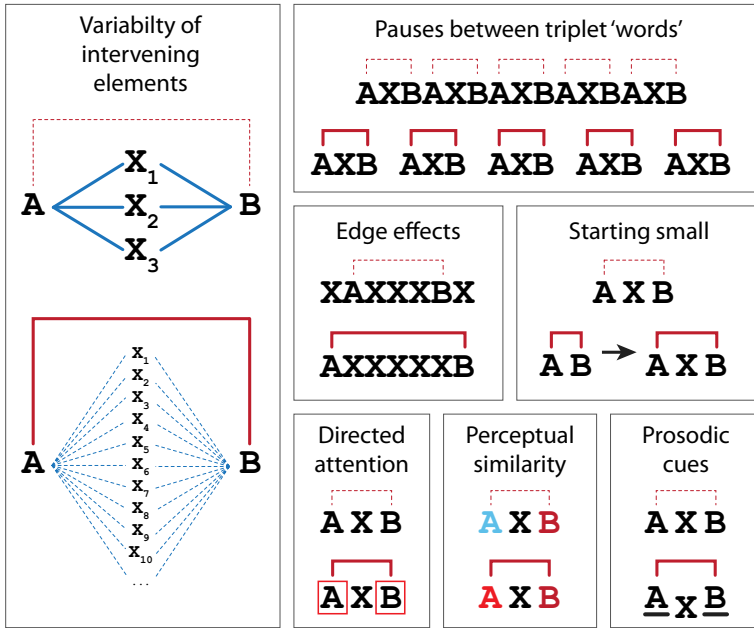


Fig. 3. Different cues aiding nonadjacent dependency learning. The learning of nonadjacent dependencies can be improved in a number of ways. These include introducing additional variability into the possible intervening elements, thus emphasizing the nonadjacent dependency (e.g., Gómez, 2002); adding pauses within streams of stimuli to denote “word” boundaries and the nonadjacent dependencies within them (e.g., Pena et al., 2002); positioning dependent stimuli on the periphery of sequences (e.g., Endress et al., 2009); initially learning the dependency between adjacent items, before introducing intervening elements (e.g., Lany & Gómez, 2008); directing attention toward dependent elements (e.g., Pacton & Perruchet, 2008); using perceptually similar dependent elements (e.g., Newport & Aslin, 2004); or the addition of prosodic cues that differentiate the dependent elements from the intervening stimuli (e.g., Grama et al., 2016). All of these different cues emphasize the relationships between nonadjacent elements and facilitate learning.

3.1.2. *Perceptual similarity between dependent stimuli*

A key cue which appears to facilitate the learning of nonadjacent dependencies is perceptual similarity between the “A” and “B” elements that differentiate them from the intervening “X” elements (Fig. 3; Creel, Newport, & Aslin, 2004; Newport & Aslin, 2004; Onnis, Monaghan, Richmond, & Chater, 2005). For example, in Gómez (2002), the “A” and “B” elements were represented by monosyllabic nonsense words, and the “X” items were all disyllabic, potentially aiding the learning of the nonadjacent dependency (Gómez, 2002). Similarly, Newport and Aslin (2004) showed that adults could learn dependencies between similar nonadjacent segments (e.g., between acoustically similar consonants separated by vowel sounds) but could not learn nonadjacent dependencies between arbitrarily related syllables (e.g., where initial “ba” predicted final “te”). When no perceptual similarities are present between the “A” and “B” stimuli, adjacent dependencies are easily learned, while there is often no evidence of nonadjacent dependency learning under these conditions (Gebhart et al., 2009; Onnis et al., 2005). Relatedly, the

addition of prosodic cues emphasizing the relationship between the “A” and “B” elements aids the learning of nonadjacent dependencies (Grama, Kerkhoff, & Wijnen, 2016). These studies suggest phonological or perceptual similarity acts as a cue facilitating the perceptual grouping of these nonadjacent elements and thus the learning of these dependencies (Newport & Aslin, 2004).

3.1.3. *Edge effects*

An additional cue that might draw attention to nonadjacent elements is positioning them at the start and the end of a sequence, as in “AXB” (Fig. 3). For example, Pena et al. (2002) presented adult participants with a continuous stream of speech stimuli grouped into trisyllabic “words” containing nonadjacent dependencies between the first and last syllables. They showed that these nonadjacent dependencies could only be learned if a brief 25 ms pause was added between each triplet “word,” thus emphasizing the word boundaries (Marchetto & Bonatti, 2013; Pena et al., 2002); but also see Onnis et al. (2005). When combined with prosodic information, pauses have also been shown to aid the learning of center embedded structures, which include nonadjacent dependencies (Fig. 1vi), within a speech stream (Mueller, Bahlmann, & Friederici, 2010). Endress and Mehler (2009) tested the importance of these “edge effects” directly, presenting adult participants with sequences in which elements “A” and “B” occurred either at the edges of a string (e.g., “AXYZB”) or within a sequence (e.g., “XAYBZ”). They found that while participants were sensitive to the reversal of the positions of the “A” and “B” elements (e.g., “BXYZA”) in either condition, they were only sensitive to the dependency between specific “A” and “B” elements when they occurred at the edges of the sequences.

Wang, Zevin, and Mintz (2017) showed that nonadjacent dependencies can be learned even when embedded in a continuous sequence of words, with no pauses or other signal-driven edge cues. They first exposed participants to a continuous natural language stream in which there was a sentence boundary every four words. This entrained subjects to parse subsequent material into periodic subsequences. Seamlessly following this pre-exposure, subjects heard a continuous stream of an artificial language containing nonadjacent “A_B” dependencies. Participants learned the nonadjacent dependency when the “A” and “B” elements appeared within a subsequence, but not if the entrained segmentation placed the dependent elements in different subsequences. Thus, top-down cues (i.e., the entrained rhythm) can facilitate the detection of nonadjacent dependencies, in the absence of surface-level cues such as pauses.

3.1.4. *Attentional effects*

The presence of pauses between triplets of syllables, dependent elements appearing at peripheral positions within a string, or linguistic entrainment appear to direct attention toward the dependent elements, thus facilitating the learning of nonadjacent dependencies. Pacton and Perruchet (2008) found that when adult participants’ attention was actively directed towards either adjacent or nonadjacent elements within a numerical sequence (by asking them to perform mathematical operations on adjacent or nonadjacent pairs of numbers), only dependencies between attended stimuli were learned. When the same task was

performed with no attentional requirements, participants implicitly learned relationships between adjacent items but failed to notice the nonadjacent dependencies (Pacton et al., 2015). These studies, along with others (e.g., de Diego-Balaguer, Martinez-Alvarez, & Pons, 2016; Friederici, Mueller, Sehm, & Ragert, 2013; Toro, Sinnott, & Soto-Faraco, 2011), support the notion that attention facilitates the learning of nonadjacent dependencies.

3.1.5. *Incremental increases in complexity: Starting small*

Nonadjacent dependency learning can be facilitated by initially learning the dependencies between items in adjacent sequence positions, before adding intervening elements. For example, 12-month-old infants were shown to be sensitive to nonadjacent dependencies, but only when they were first exposed to the same “A” and “B” items in adjacent positions in a sequence (Lany & Gómez, 2008), before testing with nonadjacent “AXB” sequences. Similarly, hierarchical relationships, which include nonadjacent dependencies (Fig. 1vi), have been shown to be learned better by adult participants when they are initially presented as pairs of adjacent items (Lai & Poletiek, 2011; Rey, Perruchet, & Fagot, 2012). These studies suggest that part of the challenge of learning nonadjacent dependencies might stem not from memory demands of holding the initial element in memory, but from detecting the relationship between temporally separated stimuli in the first place.

Taken together, this research suggests that human adults and infants are able to learn dependencies between nonadjacent elements, but that this learning strongly benefits from additional cues that highlight these elements or the relationships between them. These cues emphasize the dependencies by bringing the nonadjacent elements together—physically, perceptually, or attentionally (see Fig. 3).

While nonadjacent dependency learning is required for language learning and processing, it also applies to nonlinguistic material (e.g., Creel et al., 2004; Endress, 2010; Endress & Wood, 2011; Gebhart et al., 2009; Pacton & Perruchet, 2008), suggesting that this ability might not be domain-specific or restricted to language. In the next section, we will discuss evidence of nonadjacent dependency learning in nonhuman animals, how these abilities relate to those observed in human adults and infants, and how this might inform us about the evolutionary origins of these abilities.

3.2. *Nonhuman animals*

A number of studies have demonstrated that many nonhuman animal species can learn adjacent relationships (for recent reviews, see Santolin & Saffran, 2017; Wilson, Marslen-Wilson, & Petkov, 2017; ten Cate, 2018). However, there is also evidence that some nonhuman animals are able to learn nonadjacent dependencies in certain situations.

Nonadjacent dependency learning based on identity relations across a range of stimuli (i.e., “AXA” in Fig. 2i) has been demonstrated in squirrel monkeys (Ravignani et al., 2013) and chimpanzees (Ravignani & Sonnweber, 2017; Sonnweber et al., 2015). These studies suggest that some animals are able to detect nonadjacent dependencies, at least

between identical items. Several studies have used sequences of the form “ABA,” in which the first and last element of a sequence are identical (Murphy, Mondragon, & Murphy, 2008; Spierings & ten Cate, 2016). However, these experiments typically assess the ability to differentiate between “ABA” sequences and those of a different form (e.g., “ABB” or “AAB”), which can be identified on the basis of adjacent repetitions and do not provide evidence for learning a nonadjacent dependency.

A number of studies have tested whether nonhuman animals are able to learn nonadjacent dependencies between two different stimuli, using grammars of the form “AXB” (Fig. 2ii). However, as in humans, these studies have produced somewhat mixed results. Newport et al. (2004) conducted a study in tamarin monkeys based on a similar human experiment (Newport & Aslin, 2004). The human participants failed to learn nonadjacent dependencies between different syllables but were able to detect dependencies over both vowels and consonants (see above, Newport & Aslin, 2004). By contrast, tamarins were able to learn the dependencies based on syllables and vowels, but not consonants, suggesting that the vowel sounds (including within syllables) might be particularly salient to the monkeys (Newport et al., 2004). Recently, nonadjacent dependency learning has been demonstrated in the visual modality in tamarins (Versace et al., 2017) and baboons (Malassis, Rey, & Fagot, 2018), suggesting that these abilities are not limited to the auditory domain. de la Mora and Toro (2013) presented rats with sequences of nonsense words of alternating consonants and vowels of the form “CVCVCV,” containing nonadjacent dependencies between either the vowels or the consonants. Rats detected the dependencies in both cases (de la Mora & Toro, 2013, although see Toro & Trobalón, 2005). These studies demonstrate that at least some nonhuman animals appear to be sensitive to these types of nonadjacent dependencies, but also point to potential cross-species differences, including between humans and nonhuman animals, in how they might be learned.

Two recent studies using mixed complexity artificial grammars, which contain both adjacent and nonadjacent dependencies, showed that macaque monkeys were able to learn relationships between adjacent stimuli in the auditory or visual modality, but they found no evidence that they were sensitive to the nonadjacent dependencies (Milne et al., 2017; Wilson et al., 2015). Although humans have been reported to simultaneously learn both adjacent and nonadjacent dependencies within a mixed-complexity grammar (Romberg & Saffran, 2013a; Wang & Mintz, 2018; Wilson et al., 2015) (although not in Milne et al., 2017), it is likely that in these studies the presence of salient adjacent relationships prevented the monkeys from learning these nonadjacent dependencies (as in infants in Gómez, 2002). However, a recent EEG experiment (Milne et al., 2016) reported that violations of nonadjacent dependencies evoked similar brain potentials in macaques as had previously been reported in humans using identical stimuli (Mueller et al., 2012). These results suggest that like humans, monkeys might be sensitive to nonadjacent dependencies in some conditions.

Chimpanzees’ abilities to learn nonadjacent dependencies between visual stimuli have been assessed using operant training tasks. Some chimpanzees were able to learn nonadjacent dependencies between stimuli at the start and end of a sequence over variable distances (Fig. 1ii) and based on identity relations (Fig. 2i) or arbitrary associations

(Fig. 2ii, Sonnweber et al., 2015). Moreover, the presentation of structurally incongruous auditory stimuli appears to interfere with learning in the visual modality (Ravignani & Sonnweber, 2017). These studies suggest that chimpanzees are able to learn nonadjacent dependencies between stimuli at least at the edge of visual sequences. Endress et al. (2010) specifically assessed the salience of these “edge effects,” asking whether chimpanzees and humans could learn nonadjacent dependencies between auditory stimuli within a sequence of distracting elements. They showed that while both species were sensitive to positional effects (an “A” or “B” element occurring in an unexpected sequence position), neither humans nor chimps learned the dependency between “A” and “B” elements, embedded within a sequence (Endress et al., 2010). Finally, Chen and ten Cate (2017) used a “starting small” approach (Fig. 3) in which zebra finches first learned the dependency between two adjacent sequence elements (as in Lany & Gómez, 2008). They demonstrated that, after training with stimuli of incrementally increasing complexity, the birds were capable of detecting nonadjacent dependencies both over varying distances and at different positions within the sequences.

Taken together, these studies demonstrate that the ability to learn different types of nonadjacent dependencies (Figs. 1 and 2) is not unique to humans but may be shared by some nonhuman animals. In some cases (e.g., nonhuman primates) this might suggest a common evolutionary origin of these abilities, while in other cases (e.g., songbirds) convergent evolution might have led to the independent emergence of impressive learning abilities. Evidence of similarities and differences in sequence learning across different species allows phylogenetic analysis of the evolution of these abilities, and it may further our understanding of the evolutionary origins of cognitive processes that might represent prerequisites for language processing. Moreover, if similar abilities are indeed present across species, this will allow these processes to be studied at a neurobiological level, providing insights into the mechanisms and neural computations that underpin the learning of these types of dependencies, as has been done with adjacent dependencies (Kikuchi et al., 2017; Lu & Vicario, 2014).

4. Conclusions

The learning of dependencies between adjacent stimuli is a fundamental cognitive ability, widely conserved in the animal kingdom. However, learning and processing nonadjacent dependencies, which is required in language, appears to be more difficult. Learning is aided by the addition of cues which highlight the relationship between the nonadjacent elements, bringing them closer together either physically (e.g., Lany & Gómez, 2008), attentionally (e.g., Pacton & Perruchet, 2008), or perceptually, via acoustic similarity (Newport & Aslin, 2004), prosodic cues (Grama et al., 2016), pauses between words (Pena et al., 2002), edge effects (Endress et al., 2009; Wang et al., 2017) or by making the intervening adjacent relationships less salient (Gómez, 2002). When sufficient cues are provided, nonadjacent dependency learning has been demonstrated in human adults and infants, and some nonhuman animals, suggesting that in both evolutionary and

ontogenetic terms this ability appears to arise before the origins of language. However, some differences in learning are evident between human adults, and nonhuman animals (e.g., de la Mora & Toro, 2013; Newport & Aslin, 2004; Newport et al., 2004; Wilson et al., 2015), and further research will be required to determine how similarly these groups learn and process different forms of nonadjacent dependencies. Better understanding similarities and differences across these groups will benefit from carefully considering types of stimuli and dependencies involved (Figs. 1 and 2), and the perceptual cues available (Fig. 3). In this way, artificial grammar learning paradigms and sequence processing tasks offer great potential to explore both how language is learned over development and, via comparative studies, how it may have evolved.

Acknowledgments

We are grateful to Carel ten Cate, Judit Gervain, Clara Levelt, Chris Petkov, and Willem Zuidema for organizing the workshop on The Comparative Biology of Language Learning, which led to this article, and the Lorentz Centre for hosting the workshop.

Funding

This work was supported by a Wellcome Trust Sir Henry Wellcome Fellowship to BW (WT110198/Z/15/Z); FWO Pegasus² Marie-Curie fellowship (12N5517N) to A. Ravignani; the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 681942) to KS; DFG Forschergruppe 2253, TP1 to AK; the Netherlands Organisation for Scientific Research NWO (360-70-270) to FW; and the Chunked ANR-project (#ANR-17-CE28-0013-02) to A. Rey.

References

- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.
- Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes*, *1*(4), 249–262.
- Bahlmann, J., Schubotz, R. I., & Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *NeuroImage*, *42*(2), 525–534.
- Chen, J., & ten Cate, C. (2015). Zebra finches can use positional and transitional cues to distinguish vocal element strings. *Behavioural Processes*, *117*, 29–34.
- Chen, J., & ten Cate, C. (2017). Bridging the gap: Learning of acoustic nonadjacent dependencies by a songbird. *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*(3), 295.
- Citron, F. M., Oberecker, R., Friederici, A. D., & Mueller, J. L. (2011). Mass counts: ERP correlates of nonadjacent dependency learning under different exposure conditions. *Neuroscience Letters*, *487*(3), 282–286.

- 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(5), 1119.
- de Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology*, 7, 44.
- de la Mora, M., & Toro, J. M. (2013). Rule learning over consonants and vowels in a non-human animal. *Cognition*, 126(2), 307–312.
- De Vries, M., Christiansen, M. H., & Petersson, K. M. (2011). Learning recursion: Multiple nested and crossed dependencies. *Biolinguistics*, 5(1/2), 010–035.
- De Vries, M. H., Petersson, K. M., Geukes, S., Zwitserlood, P., & Christiansen, M. H. (2012). Processing multiple non-adjacent dependencies: Evidence from sequence learning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1598), 2065–2076.
- Demuth, K., & Weschler, S. (2012). The acquisition of Sesotho nominal agreement. *Morphology*, 22(1), 67–88.
- Endress, A. D. (2010). Learning melodies from non-adjacent tones. *Acta Psychologica*, 135(2), 182–190.
- Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, 105(2), 247–299.
- Endress, A. D., Carden, S., Versace, E., & Hauser, M. D. (2010). The apes' edge: Positional learning in chimpanzees and humans. *Animal Cognition*, 13(3), 483–495.
- Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, 60(3), 351–367.
- Endress, A. D., Nespore, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, 13(8), 348–353.
- Endress, A. D., & Wood, J. N. (2011). From movements to actions: Two mechanisms for learning action sequences. *Cognitive Psychology*, 63(3), 141–171.
- Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303(5656), 377–380.
- Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, USA*, 103(7), 2458–2463.
- Friederici, A. D., Mueller, J. L., & Oberecker, R. (2011). Precursors to natural grammar learning: Preliminary evidence from 4-month-old infants. *PLoS ONE*, 6(3), e17920.
- Friederici, A. D., Mueller, J. L., Sehm, B., & Ragert, P. (2013). Language learning without control: The role of the PFC. *Journal of Cognitive Neuroscience*, 25(5), 814–821.
- Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, 147, 70–74.
- Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, 16(3), 486–490.
- Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition*, 98(3), B67–B74.
- Gervain, J., & Werker, J. F. (2013). Learning non-adjacent regularities at age 0; 7. *Journal of Child Language*, 40(4), 860–872.
- Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431–436.
- Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7(2), 183–206.
- Grama, I. C., Kerkhoff, A., & Wijnen, F. (2016). Gleaning structure from sound: The role of prosodic contrast in learning non-adjacent dependencies. *Journal of Psycholinguistic Research*, 45(6), 1427–1449.
- Grama, I., Wijnen, F., & Kerkhoff, A. (2013). Constraints on non-adjacent dependency-learning: Distance matters an artificial grammar learning study with adults. Paper presented at the Boston University Conference on Language Development 37 online proceedings supplement, Boston, MA.

- Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (2014). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in Psychology, 5*(175), 1–10.
- Kerkhoff, A., De Bree, E., De Klerk, M., & Wijnen, F. (2013). Non-adjacent dependency learning in infants at familial risk of dyslexia. *Journal of Child Language, 40*(1), 11–28.
- Kikuchi, Y., Attaheri, A., Wilson, B., Rhone, A. E., Nourski, K. V., Gander, P. E., Kovach, C. K., Kawasaki, H., Griffiths, T. D., & Howard III, M. A. (2017). Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS Biology, 15*(4), e2000219.
- Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., & Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Developmental Science, 14*(5), 1100–1106.
- Lai, J., & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition, 118*(2), 265–273.
- Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science, 19*(12), 1247–1252.
- Lu, K., & Vicario, D. S. (2014). Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proceedings of the National Academy of Sciences, USA, 111*(40), 14553–14558.
- Malassis, R., Rey, A., & Fagot, J. (2018). Non-adjacent dependencies processing in human and non-human primates. *Cognitive Science, 42*, 1677–1699.
- Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language acquisition. *Cognitive Psychology, 67*(3), 130–150.
- Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77–80.
- Milne, A. E., Mueller, J. L., Männel, C., Attaheri, A., Friederici, A. D., & Petkov, C. I. (2016). Evolutionary origins of non-adjacent sequence processing in primate brain potentials. *Scientific Reports, 6*, 36259.
- Milne, A. E. et al. (2017). Auditory and visual sequence learning in humans and monkeys using an artificial grammar learning paradigm. *Neuroscience*, <http://doi.org/10.1016/j.neuroscience.2017.06.059>
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology, 1*, 31.
- Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2010). Learnability of embedded syntactic structures depends on prosodic cues. *Cognitive Science, 34*(2), 338–349.
- Mueller, J. L., Friederici, A., & Mannel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences, 109*(39), 15953–15958.
- Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure—An ERP study in adult learners. *BMC Neuroscience, 10*(1), 89.
- Murphy, R. A., Mondragon, E., & Murphy, V. A. (2008). Rule learning by rats. *Science, 319*(5871), 1849–1851.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127–162.
- 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. *Cognitive Psychology, 49*(2), 85–117.
- Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language, 53*(2), 225–237.
- Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 80.
- Pacton, S., Sobaco, A., & Perruchet, P. (2015). Is an attention-based associative account of adjacent and nonadjacent dependency learning valid? *Acta Psychologica, 157*, 195–199.
- Pena, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*(5593), 604–607.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233–238.

- Petersson, K. M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, *120*(2), 83–95.
- Ravignani, A., & Sonnweber, R. (2017). Chimpanzees process structural isomorphisms across sensory modalities. *Cognition*, *161*, 74–79.
- Ravignani, A., Sonnweber, R.-S., Stobbe, N., & Fitch, W. T. (2013). Action at a distance: Dependency sensitivity in a New World primate. *Biology Letters*, *9*(6), 20130852.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour*, *6*(6), 855–863.
- Rey, A., Perruchet, P., & Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: Evidence from baboons (*Papio Papio*). *Cognition*, *123*(1), 180–184.
- Romberg, A. R., & Saffran, J. (2013a). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science*, *37*, 1290–1318.
- Romberg, A. R., & Saffran, J. R. (2013b). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science*, *37*(7), 1290–1320.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 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*(1), 27–52.
- Santolin, C., & Saffran, J. R. (2017). Constraints on statistical learning across species. *Trends in Cognitive Sciences*, *22*(1), 52–63.
- Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition*, *18*, 733–745.
- Spierings, M. J., & ten Cate, C. (2016). Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *Proceedings of the National Academy of Sciences, USA*, *113*(27), E3977–E3984.
- Teinonen, T., Fellman, V., Nääätänen, R., Alku, P., & Huutilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*(1), 21.
- ten Cate, C. (2018). The comparative study of grammar learning mechanisms: Birds as models. *Current Opinion in Behavioral Sciences*, *21*, 13–18.
- Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2011). Generalizing linguistic structures under high attention demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 493.
- Toro, J., & Trobalón, J. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*(5), 867–875.
- van den Bos, E., Christiansen, M. H., & Misyak, J. B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*, *67*(4), 507–520.
- Versace, E., Rogge, J. R., Shelton-May, N., & Ravignani, A. (2017). Positional encoding in cotton-top tamarins (*Saguinus oedipus*). *bioRxiv*, 186692.
- Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2016). Concurrent statistical learning of adjacent and nonadjacent dependencies. *Language Learning*, *66*(1), 8–30.
- Wang, F. H., & Mintz, T. H. (2018). Learning nonadjacent dependencies embedded in sentences of an artificial language: When learning breaks down. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 604.
- Wang, F. H., Zevin, J. D., & Mintz, T. H. (2017). Top-down structure influences learning of nonadjacent dependencies in an artificial language. *Journal of Experimental Psychology: General*, *146*(12), 1738.
- Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved sequence processing in primate frontal cortex. *Trends in Neurosciences*, *40*, 72–82.
- Wilson, B., Smith, K., & Petkov, C. I. (2015). Mixed-complexity artificial grammar learning in humans and macaque monkeys: Evaluating learning strategies. *European Journal of Neuroscience*, *41*(5), 568–578.