Artificial Grammar Learning in Primates: Behaviour and Neuroimaging

Benjamin Wilson

Submitted for the degree of Doctor of Philosophy

Institute of Neuroscience,

Newcastle University

September 2013

General Abstract

Neuroimaging studies have shown that natural language processes engage left hemisphere perisylvian brain regions. Artificial Grammars (AG), which are designed to emulate aspects of language syntactic structure, recruit comparable brain areas. Nonhuman animals have been shown to learn a range of different AGs. However, no data is currently available regarding the brain areas that support these processes. In this thesis, I combined behavioural artificial grammar learning (AGL) and fMRI experiments to generate insights regarding language evolution, and as a first step to developing animal model systems for aspects of language processing. These experiments provide novel evidence that nonhuman primates are able to learn a non-deterministic AG, designed to emulate some of the variability of the structure of sentences in natural language, and demonstrated notable correspondences between the brain regions involved in macaque and human AGL. I developed a quantitative method to compare AGL abilities across species and studies, and a novel eye-tracking technique with which to collect objective behavioural data. Using this technique, and a refined version of a traditional video-coding paradigm, I demonstrated that Rhesus macaques notice violations of the AG structure and that these results could not be explained by reliance on simple cues. Common marmosets also showed evidence of AGL however, these results may have been driven by simple learning strategies. Comparative fMRI experiments showed that, in humans, violations of the AG activated a number of perisylvian brain regions associated with language processing, including the ventral frontal cortex (vFC), temporal and temporo-parietal regions, although not Broca's area (BA44/45). In Rhesus macaques, comparable patterns of activation were seen in the ventral frontal cortex and temporo-parietal regions. Additional activation in BA44/45 in macaques provides interesting insights into the evolution of this region. These experiments provide novel evidence regarding the AGL capabilities of nonhuman primates, and the brain areas that support them, suggesting that some language related functions may represent generic, rather than language specific processes. Therefore, some of the brain regions involved in AGL in both species might share a common evolutionary heritage, and therefore Rhesus macaques might represent a valuable animal model system for aspects of language processing.

Acknowledgements

I would like to thank all of the people who have helped and supported me throughout my time at Newcastle University.

First and foremost, I would like to thank my supervisor, Dr Christopher Petkov, without whose unwavering support and guidance I would never have attempted, let alone completed, such an ambitious project. I will always remember his motivation, dedication, scientific drive and integrity, and when things got tough, always reminding me that you have to fight for your data.

I would also like to thank the rest of the Laboratory of Comparative Neuropsychology, past and present. Members of the lab have helped code countless hours of videos of primate behaviour; discussed, debated and criticised proposed experiments and analyses; and provided invaluable feedback on drafts of papers and articles. For all this, and much more, I would like to thank Heather Slater, Alice Milne, Adam Attaheri, Dr Yukiko Kikuchi, Hannah Bassirat, Matthew Collison, Harriet Wright and James Jardine.

I am also grateful to Prof Adrian Rees, Dr Tom Smulders, Prof Tim Griffiths, Dr Kenny Smith and Prof William Marslen-Wilson, who have helped guide and shape the direction of this research with insightful discussion, comments and suggestions.

I would also like to thank everyone who has helped me learn the wide range of skills and techniques involved in this research. For expert guidance and training in all things primate related, I would like to thank Mike White, Ashley Waddle, Laura Watson, Lee Reed and Stevie O'Keefe, as well as Prof Paul Flecknell and Dr Aurelie Thomas for veterinarian support. For fMRI support I would like to thank Louise Morris, Carol Smith, Tim Hodgson, Dr Li Sun and particularly Dr David Hunter. I also thank Vince Wiley and Charlie Gray for machine shop support.

Thanks must also go to my friends and family for discussion, support and friendship during both the good times and the stressful ones, particularly my parents, Kay and Steve; my brother, Joe; and my friends, Willemijn, Katja, Gillian, Steph, and of course, all the people already mentioned, who I view not only as colleagues, but also as friends.

Finally, I must thank all of the participants, and in particular the primates who have contributed to this work: Carl, Paul, Chris and Watson; Andy, Anne, Charlie and Eddie, without whom none of this research would have been possible.

Table of Contents

List of Figures

List of Tables

Abbreviations

AC	Auditory Cortex
AF	Arcuate Fasciculus
AG	Artificial Grammar
AGL	Artificial Grammar Learning
BA	Brodmann Area
BOLD	Blood Oxygen Level Dependent
CFG	Context Free Grammar
CI	Confidence Interval
CV	Consonant-Vowel
DTI	Diffusion Tensor Imaging
EEG	Electroencephalography
EFCS	Extreme Fibre Capsule System
EPI	Echo Planar Imaging
FEAT	FMRI Expert Analysis Tool
FLH	Formal Language Hierarchy
fMRI	functional Magnetic Resonance Imaging
FMRIB	Oxford Centre for Functional MRI of the Brain
FOP	Frontal Operculum
FSG	Finite State Grammar
FSL	FMRIB Software Library
GE-EPI	Gradient Echo Echo-Planar Imaging
GLM	General Linear Model
IFG	Inferior Frontal Gyrus
LCA	Last Common Ancestor
LSD	Least Significant Difference test
MDEFT	Modified Driven Equilibrium Fourier Transform
MEG	Magnetoencephalography
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
MYA	Million Years Ago
PET	Positron Emission Tomography
RF	Radio Frequency
RM-ANOVA	Repeated Measures Analysis Of Variance
RMS	Root Mean Squared
ROI	Region Of Interest
rTMS	repetitive Trascranial Magnetic Stimulation
SD	Standard Deviation
SEM	Standard Error of the Mean
SLF	Superior Longitudinal Fasciculus
tDCS	transcranial Direct Current Stimulation
TE	Echo Time
TP	Transitional Probability
TR	Repetition Time
UF	Uncinate Fasciculus
vFC	ventral Frontal Cortex
WM	Working Memory

Chapter 1: Introduction

Language is a defining human trait. It is ubiquitous across all human cultures, but unique in both scale and complexity within the animal kingdom (Hauser *et al.*, 2002; Fitch, 2010). Semantic representations allow the communication of abstract concepts and combinatorial syntax allows the generation and transmission of an infinite number of ideas from a limited vocabulary (Pinker, 1994). Language is not only critical to human society, but may even have been a necessary requirement for the evolution of the higher brain functions which characterise our species (Hauser *et al.*, 2002; Christiansen & Kirby, 2003b; Bickerton, 2009). As such, human language has received a vast amount of research from a wide range of scientific disciplines, from Linguistics (e.g., Chomsky, 1957) to Psychology (e.g., Pinker & Bloom, 1990) and more recently Neurobiology (e.g., Friederici, 2011). Such research has led to dramatic growth in our understanding the structure of natural languages and how they are learned and processed. However, despite the successes of this work, it has yet to elucidate how such incredible linguistic abilities may have evolved in humans, but not in other species.

In 1866, frustrated by the speculative nature of discussions, the Linguistic Society of Paris banned all discussion of the subject of language evolution (Christiansen & Kirby, 2003b; Newmeyer, 2003; Hauser *et al.*, 2007). Despite a recent revival in interest in the topic (Christiansen & Kirby, 2003b; 2003a; Fitch, 2010; Hurford, 2012), the evolutionary origins of language are still unclear, and has even been described as 'the hardest problem in science' (Christiansen & Kirby, 2003b). While this contentious claim may be debated, there are certainly serious challenges to explaining how human language abilities may have evolved. Physical evidence for the existence of language may be derived from archaeological records of early writing, but the oldest of these date back to less than 10000 years ago, a blink of the eye in evolutionary terms (Houston, 2008). Furthermore, while the production and perception of language relies to some extent on physiological adaptations (for example articulatory control, a descended larynx, good auditory discrimination, Fitch, 2000), the ultimate origin of all linguistic ability is the brain. While the fossil record may provide interesting evidence about gross changes in brain size over evolutionary time (Dunbar, 1998; Fitch, 2000), this information is unable to address what linguistic or proto-linguistic abilities may have been possessed by our now extinct ancestors.

An alternative approach to the study of language evolution is to consider the abilities of closely related species. If some of the cognitive abilities that are required for language arose from pre-existing, non-linguistic mechanisms in our ancestral past, it is possible that such abilities may still be present in our extant primate cousins (Hauser *et al.*, 2002; Petkov & Wilson, 2012). However, this approach has received significant scepticism. A number of notable scientists have argued that language developed so recently in our evolutionary history, that no traces would be present in even our closest evolutionary relatives (Pinker & Bloom, 1990). It is possible that all aspects of human language may postdate human-chimpanzee speciation (approximately 5 million years ago, Sun *et al.*, 2012); in which case no traces of language related abilities may be present in any nonhuman species. If this hypothesis were to prove accurate, then no primates would possess even the rudimentary cognitive mechanisms on which humans rely for language learning and processing. Furthermore, if the animals were to show behavioural similarities to humans in language related tasks, it would be likely that the brains of these species would process these computations in fundamentally different ways.

Alternatively, human language capabilities may have more distant evolutionary roots, building on brain networks and cognitive abilities (not necessarily related to language Hauser *et al.*, 2002) shared by our common ancestors with extant primates. In this case, the study of nonhuman animals, particularly closely related primate species, could provide invaluable insights into the evolution of human language abilities (Fitch, 2000; Petkov & Wilson, 2012). Therefore, in order to even identify how useful a comparative approach to language evolution may be, it is necessary to explore what abilities may be shared between related primate species.

In this chapter I will argue that taking a comparative approach may provide unique insights into how language may have evolved. I will discuss how the combination of both behavioural testing and neuroscientific techniques present not only the potential to inform us about language evolution, but also the opportunity to develop animal model systems in which aspects of language learning and processing may be tested at a cell and molecular level.

1.1. Studying language related abilities in nonhuman animals

The genesis of our linguistic abilities is currently unclear. A great deal of debate surrounds the discussion of language evolution, but the majority of such work focuses on systems of communication (vocal or otherwise, e.g. Corballis, 2003), much more recent than anything shared by our extant primate cousins. A number of notable linguists argue that language abilities may have evolved rapidly as a result of increasing brain size in the human lineage (e.g., Chomsky, 1972; Gould & Lewontin, 1979). Alternatively, many psychologists have suggested that linguistic ability may be a desirable trait, either increasing an individual's survival or reproductive success, and therefore would spread and evolve as a result of natural or sexual selection (Pinker & Bloom, 1990; Deacon, 1997; Miller, 2001). While these hypotheses are appealing, they fail to explain the earliest origins of language; before language could be used to usefully communicate meaningful information, our ancestors must have learned to combine simple calls in meaningful ways.

To investigate the initial origins of the abilities which led to language in humans, but not other primate species, it may be necessary to consider our more distant ancestors. It is important to note that the abilities of nonhuman animals that might have evolved into the human language system need not have formed part of the animals' communication systems, as other brain networks and cognitive mechanisms may have been co-opted into a new function (Hauser *et al.*, 2002; Fitch & Hauser, 2004). These abilities, which are required for language but are not specific to it, have been referred to as "the language faculty in the broad sense" (Hauser *et al.*, 2002). To assess whether any extant species do possess such abilities, it is necessary to explore both the simple, fundamental cognitive mechanisms required for language in humans, and the abilities of nonhuman animals to perform similar tasks. Furthermore, to convincingly demonstrate that the traits observed in nonhuman animals are related to any sort of human linguistic ability, it would be valuable to demonstrate that both species recruit similar networks of brain regions to perform such tasks. Alternatively, we may be forced to conclude that while a nonhuman primate may be able to perform a task which may appear related to language in humans (and may even recruit the language network in the human brain) that they do so using very different mechanisms which are unlikely to have evolved into the system seen in humans.

Do any extant nonhuman species possess any abilities which may share an evolutionary heritage with our own language abilities? Is there any value in taking a comparative approach to studying language evolution? In the following section I will discuss the evidence relating the abilities of nonhuman animals to language in humans.

1.1.1. Animal communication abilities

While human language is uniquely complex, many species rely on vocal communication for both survival and reproduction. Most species recognise conspecific vocalisations compared to other sounds or the calls of other animals (Zoloth *et al.*, 1979; Seyfarth *et al.*, 1980) and easily identify specific individuals (Gentner & Hulse, 1998; Rendall *et al.*, 1998; Fitch & Fritz, 2006; Ghazanfar *et al.*, 2007). Beyond this apparently common ability to recognise the calls of familiar or related animals, many species utilise more complex calls in response to specific events, objects or individuals. For example, the males of a wide range of species produce mating calls, ranging from very simple vocalisations primarily advertising the location and availability of a male, (elephant seals, Bartholomew & Collias, 1962; deer, McComb, 1991; Rhesus macaques, Hauser, 1993), to highly elaborate songs in more vocally adept species (Okanoya, 2004a). Many animals also produce calls in response to the presence of food (Snowdon *et al.*, 1997; Clay *et al.*, 2012), in some cases including specific vocalisations upon encountering highly desirable foods (Rhesus macaques: Hauser & Marler, 1993a; chimpanzees: Slocombe & Zuberbuhler, 2005; Bonobos: Clay & Zuberbuhler, 2009). Similarly, alarm calls are common in social species to warn conspecifics of predation risks (Struhsaker, 1967; Seyfarth *et al.*, 1980; Hauser & Marler, 1993b; 1993a; Zuberbuhler, 2000d; 2000a), in some cases with even specifying the type of predatory threat (Seyfarth *et al.*, 1980; Arnold & Zuberbuhler, 2006; Ouattara *et al.*, 2009b; 2009a). Social interactions often rely upon vocal communication, for example many monkeys use specific calls while attempting to elicit group movement or to instigate affiliative, aggressive or sexual interaction (Seyfarth *et al.*, 1980; Seyfarth & Cheney, 1984; Hauser & Marler, 1993b; 1993a; Harcourt & Stewart, 1996; Seyfarth & Cheney, 1999; for reviews, see Fitch, 2000; Hauser *et al.*, 2002; Lemasson *et al.*, 2004; Lemasson & Hausberger, 2011).

These vocalisations allow the communication of crucial information to conspecifics (and in some cases to members of other species, Zuberbuhler, 2000a; Seiler *et al.*, 2013). However, these calls typically refer to a single specific object, individual or action. Human language is much more complex than this. It allows words or phrases to be combined in meaningful ways, allowing the communication of a limitless number of thoughts, ideas, statements and questions. One of the defining qualities of human language is the incredible generative capacity afforded by the syntax of natural languages, allowing a finite vocabulary to be combined into an infinite number of meaningful sentences (Pinker, 1994). While animals vary in the number and utility of vocalisations they can produce, without a combinatorial system like the syntax of natural language their communication abilities would remain an order of magnitude simpler and less effective. In the following section I will discuss syntax in human language, and review the evidence for potentially 'proto-syntactic' combinatorial communication systems in nonhuman species.

1.2. Syntax

Syntax is fundamental to human language. Syntactic rules allow the generation of an infinite number of sentences from a finite number of words and rules, language "makes infinite use of finite means" (Humboldt, 1836). Therefore, syntactic abilities are minimally reliant upon the ability to evaluate grammatical relations between words in an utterance. The production and comprehension of language rely on both the speaker and listener applying the same syntactic rules to the sequence of words in a sentence, to arrive at a shared understanding. Based on this definition, syntax is unarguably unique in the animal word, occurring only as part of human languages. However, if we consider that the complexity of human language and linguistic ability may have evolved from simpler cognitive mechanisms in our evolutionary ancestors, it becomes valuable to consider the fundamental basis of such syntactic abilities. A key component of syntactic abilities is, therefore, the capacity to learn how expressions in a sequence are appropriately structured. It is therefore important to ask, to what extent do animals produce meaningful sequences of vocalisations?

1.2.1. Combinatorial communication systems in nonhuman animals

A huge variety of animals produce sequences of calls, however for these to constitute even the simplest form of 'proto-syntactic' communication it is necessary that the order of these vocalisations affects the response elicited in the listeners. It may be noteworthy that it is difficult to identify what the individual vocalisations may 'mean', either to the caller or recipient, but if different call combinations produce consistent, distinct behavioural responses it is likely that the information being communicated may be reliant on the order of the calls. For example, an excited dog may bark persistently, but altering the order in which these barks occur is unlikely to have any meaningful impact on any other dogs that happen to be listening. A number of species produce sequenced vocalisations in which the order of the vocalisations does appear to have an impact on the behaviour, and presumably mental state, of listening conspecifics (e.g., Okanoya, 2004b). Adept vocal communicators, for example many songbird species, learn to produce songs, specific sequences of motifs, learned from the parents and used through adulthood, primarily as part of mating displays (e.g., Okanoya, 2004b; Catchpole & Slater, 2008). The complexity and quality of these songs acts as an honest indicator of fitness (Zahavi, 1975) and corresponds strongly to the mating success of the males (Searcy & Anderson, 1986; MacDougall-Shackleton, 1997; Gil & Gahr, 2002; Catchpole & Slater, 2008).

The communication systems of these species suggest that some animals do combine vocalisations in consistent ways and that the ability to do this successfully has an observable impact on their mating success. In these animals accurate reproduction of a song appears to act as an indicator of fitness; virtuosity is a desirable trait to females, so those males who produce the best songs enjoy the highest reproductive success. Therefore, by sexual selection the ability to learn complex songs evolves in the population. However, in these species songs are typically not used to communicate information, except in the sense that they communicate an individual's ability to produce song of that quality. Such sequencing of sounds in specific orders is termed 'phonological syntax', and is distinct from the 'semantically compositional syntax' of language (Marler & Tenaza, 1977), whereby the order of words in a sentence affects the meaning of the sentence, for example the difference between 'the boy kicked the girl' and 'the girl kicked the boy', (Marler, 2000; Bickerton & Szathmary, 2009; Tallerman, 2011; Hurford, 2012).

The lack of meaningful, referential song motifs in song birds highlight a fundamental difference between human language and the phonological syntax displayed by these species. However, this distinction does not immediately imply that the study of such systems may not be highly informative. It is difficult to imagine a creature capable of semantically compositional syntax but incapable of phonological syntax; to appreciate that the order of elements in a sequence affects the meaning of that sequence it is first necessary to be able to learn and recognise such ordering. It is therefore possible that the more complex syntactic abilities of humans may have evolved from simpler, relatively rudimentary phonological sequencing abilities.

The sequencing abilities of songbirds and other vocal learning species therefore provide a valuable model to consider the evolution of phonological syntax. However, any abilities apparently shared by humans and distantly related species such as songbirds, whales, mice and other such species are likely the result of convergent evolution rather than common descent (Petkov & Jarvis, 2012). Furthermore, the vocal learning and sequencing abilities of songbirds rely on entirely different neurobiology to that of humans (Feenders *et al.*, 2008; Petkov & Jarvis, 2012). Therefore, an important question becomes, do any more closely related nonhuman primates appear to sequences their vocalisations (even phonologically) in any meaningful way?

1.2.2. Meaningful call combinations in nonhuman primates

Nonhuman primates appear to lack the complex vocal learning abilities that allow species such as songbirds to generate large numbers of song motifs from which to construct their songs (Janik & Slater, 1997; Boughman, 1998; Fitch, 2000; Noad *et al.*, 2000; Jarvis, 2004; Snowdon, 2009; Bolhuis *et al.*, 2010; Pepperberg, 2010; Petkov & Wilson, 2012). Instead, most nonhuman primates possess species specific repertoires of calls and vocalisations, which appear predictably in all members of the species (e.g. Hauser & Marler, 1993b; 1993a; Candiotti *et al.*, 2012b). While there may be some flexibility in how these calls are produced (pygmy marmosets: Elowson & Snowdon, 1994; Snowdon & Elowson, 1999; Campbell's monkeys: Lemasson & Hausberger, 2004; Japanese macaques: Koda *et al.*, 2008; baboons: Ey *et al.*, 2009; chimpanzees: Slocombe *et al.*, 2010) they typically represent only acoustical variants of a small number of main call types (Candiotti *et al.*, 2012b) and are vastly more limited than the level of flexibility afforded to true vocal learners (Petkov & Jarvis, 2012). Despite this, nonhuman primates' phylogenetic proximity to humans may suggest that they are relevant to investigating language evolution (Lemasson & Hausberger, 2011).

A number of nonhuman primate species have been reported to attempt to overcome their sparse vocabularies by concatenating some of their calls in meaningful ways. Combining calls in such a manner is a more efficient, although more cognitively demanding method for increasing the size of a species communicative repertoire (Jackendoff, 1999). For example, a number of guenons, Old World monkeys which live in dense forests, emit social contact calls to maintain group cohesion where visibility is poor (Uster & Zuberbuhler, 2001), and specific alarm calls in the presence of aerial or terrestrial predators (Zuberbuhler, 2000c; 2000a; Zuberbuhler & Jenny, 2002). For example, putty-nosed monkeys, produce distinct 'pyow' calls to indicate the presence of leopards and 'hack' calls in response to crowned eagles (Arnold & Zuberbuhler, 2006; 2008), but also combine these calls into a 'pyow-hack' composite indicating the presence of both predators. Playback experiments have demonstrated that these 'pyow-hack' calls lead to significantly increased group movement while the calls in isolation trigger reduced movement and increased vigilance. The authors argue that the combination of two calls, which have associated behavioural responses when presented independently, to produce a third call which causes a new response in listeners demonstrates semantic combinatorial abilities (Arnold & Zuberbuhler, 2006; 2008).

Campbell's monkeys similarly produce specific vocalisations in response to sightings of leopards and eagles (Zuberbuhler, 2002) and both concatenate these vocalisations in specific ways in particular social contexts (Ouattara *et al.*, 2009b) and also modify the specificity of these calls through the addition of affixes (Ouattara *et al.*, 2009a). Furthermore, Zuberbühler demonstrated that Diana monkeys, which live in overlapping territories and naturally attend to the alarm calls of the other species, respond differently to the vocalisations of the Campbell's monkeys depending whether a modifying call was concatenated with their alarm calls (Zuberbuhler, 2000b). Further evidence of nonhuman primates combining calls have been observed in many species, including chimpanzees (Crockford & Boesch, 2005), bonobos (Clay & Zuberbuhler, 2009; 2011), gibbons (Clarke *et al.*, 2006), Guereza colobus monkeys (Schel *et al.*, 2010) and Diana monkeys (Candiotti *et al.*, 2012a).

While these studies present convincing evidence that some nonhuman primates produce sequences of calls, and that in some cases the meanings of such calls affect the information communicated, it is clear that these abilities are not comparable with human language (e.g., Hauser *et al.*, 2002; Candiotti *et al.*, 2012a). The number of different sequences of vocalisations that can be produced is inherently limited by the size of an animal's 'vocabulary'. As such, nonhuman primates, with their limited vocal learning abilities (Petkov & Jarvis, 2012) and limited number of available vocalisations (Hauser & Marler, 1993b; 1993a), would be unable to produce human-like levels of communicative complexity, even if they were to possess sufficient sequencing abilities. However, a lack of complexity in the vocalisations of a species does not necessarily imply that the species does not possess the ability to recognise how a sequence is structured. It is possible that they might possess the required cognitive abilities but may not use these as part of their vocal communication (Hauser *et al.*, 2002; Petkov & Wilson, 2012). Furthermore, one might even predict animals to possess better perception than production abilities, not only due to the physiological requirements of producing complex vocalisations (Fitch, 2000), but also because the ability to mentally process combinations of calls is clearly a prerequisite to being able to produce any meaningful sequences of vocalisation (Moore, 2004; Petkov & Wilson, 2012). Therefore, the reported limitations of the vocal production abilities of nonhuman primates do not preclude the possibility that some aspects our syntactic abilities may have evolved from cognitive abilities shared with our last common ancestor with other primates. In the following section I will discuss how we may experimentally assess perception of syntax-related sequences.

1.3. Artificial Grammar Learning

Artificial Grammar Learning (AGL) paradigms aim to tap into the abilities that humans use to learn syntactically structured sequences, by testing participants' abilities to learn and process simple 'grammars' designed to emulate specific aspects of language without requiring the full range of human syntactic abilities (Reber, 1967; Marcus *et al.*, 1999; Friederici, 2004; Petersson *et al.*, 2012). Typically an Artificial Grammar (AG) is designed to model an interesting feature of language, and is then used to generate a number of 'grammatical' strings or sequences of elements, for example nonsense words or symbols, which follow this AG structure. Participants will initially be habituated to these sequences, before being tested with a combination of sequences generated by the AG and new 'ungrammatical' sequences that violate the AG structure. A wide variety of response measures can then be used based on the hypotheses or the population being tested (e.g. implicit orienting responses, explicit forced choice responses, neuroimaging experiments, Reber, 1967; Friederici *et al.*, 2006a; Saffran *et al.*, 2008). It is therefore possible to determine whether the participants are able to distinguish between the 'grammatical' and 'ungrammatical' sequences, and therefore whether they have learned the AG. These experiments require little or no instruction (depending on the method of response required from the participants) and importantly allow the assessment of participants' sequencing capabilities without requiring any of the additional abilities that may be involved in human language (e.g. attaching semantic meanings to words). This approach has been highly successful in exploring how adult humans process different linguistic structures (Reber, 1967; Bahlmann *et al.*, 2008; Petersson *et al.*, 2012) as well as testing pre-linguistic infants (Gomez & Gerken, 1999; Marcus *et al.*, 1999; Saffran *et al.*, 1999; Saffran *et al.*, 2008) and even nonhuman animals (Hauser *et al.*, 2001; Fitch & Hauser, 2004; Newport *et al.*, 2004; Gentner *et al.*, 2006; Murphy *et al.*, 2008; Saffran *et al.*, 2008; Hauser & Glynn, 2009; van Heijningen *et al.*, 2009; Abe & Watanabe, 2011; Stobbe *et al.*, 2012). AGL paradigms may offer a unique opportunity to investigate, in isolation from other potentially confounding factors, the extent to which nonhuman animals may be able to learn the structure of artificial grammars which can be related to human syntax. This approach allows us to determine what 'proto-syntactic' abilities nonhuman animals may possess, and provides the opportunity to investigate the extent to which human sequence processing capabilities may share common ancestry with extant primate species.

1.3.1. Artificial Grammar Learning in nonhuman animals

All human languages are complex, and require speech to be processed at a number of levels, potentially employing a range of different cognitive mechanisms. Speech perception requires an incoming stream of complex auditory stimuli to be accurately parsed into syllables and words. Lexical information must then be attached to these words, before phrases can be interpreted based on the syntax of the language, and ultimately the whole sentence can be comprehended (Chomsky, 1965). This is a highly demanding task, and as such maybe it is of little surprise that nonhuman animals are unable to perform it. However, to understand what abilities other species may possess, it is beneficial to initially ask to what extent nonhuman animals may have some of these abilities, which seem to be prerequisites for human language, in isolation from the others. In this way it may be possible to highlight where human abilities are truly unique, but also where our language processing capacities may show some overlap with our primate cousins.

A number of nonhuman species have been tested using a variety of AGL paradigms. In this section I will briefly review the evidence that animals can learn different AGs with reference to how these may be necessary for aspects of human syntax. In Chapter 2, I will consider how these disparate studies may be quantitatively compared to better understand what cognitive abilities may be shared by different species, with the goal of guiding further research to identify where human and nonhuman abilities may cumulate or ultimately diverge.

An initial stage in language processing is the identification of words and word boundaries within a speech stream. While phonology and stress patterns are often used to identify words within a sentence (Cutler and Carter, 1987), these acoustic cues are not always present, and may be of limited use to new language learners in identifying word boundaries (Cole *et al.,* 1980; Aslin *et al.*, 1996; Saffran *et al.*, 1996b), therefore, how do new language learners recognise word boundaries to identify words, the fundamental building blocks of language? One possibility is to consider the statistical regularities between pairs of phonemes within and across word boundaries. Phonemes within words co-occur more regularly than those between words, therefore present a valuable cue to identifying words within a speech stream. Human participants have been shown to use these transitional probabilities (TPs) to identify boundaries in continuous streams of meaningless syllables (Saffran *et al.*, 1996a; Saffran *et al.*, 1996b; Saffran *et al.*, 1997; Aslin *et al.*, 1998; Gomez & Gerken, 1999; Newport & Aslin, 2004; Newport *et al.*, 2004) or tones (Saffran *et al.*, 1999). To investigate whether this ability may be uniquely human, Hauser *et al.* (2001) were motivated to investigate whether such statistical learning abilities may also be possessed by cotton-top tamarins (*Saguinus oedipus*). Like previous studies in humans, Hauser and colleagues found that after habituation to a stream of syllables arranged into tri-syllabic nonsense words, the tamarins were able to use the transitional probabilities between phonemes to discriminate between 'words' (which were presented in the habituation stream), 'part words' (which were present in the habituation stream, but crossed a word boundary) and 'non-words' (which were not present in the habituation stream). These results suggest that while tamarins naturally show very limited sequencing of their own vocalisations (Weiss *et al.*, 2001) they appear to be implicitly sensitive to the structure of these artificial grammar sequences (Hauser *et al.*, 2001). Furthermore, the same group has demonstrated similar effects when the tri-syllabic nonsense words in the speech stream contained regularities only between the first and third syllable, therefore requiring the animals to recognise long distance, non-adjacent relationships (Newport *et al.*, 2004). These studies suggest that while such abilities are necessary for language acquisition (Saffran *et al.*, 1996b), they do not appear to be uniquely human, and are likely generic rather than language specific.

Once word boundaries in a sentence can be identified, syntactic processing requires understanding of the relationships between sounds now established as "words", and an ability to generalise these relationships to new sentences. A number of studies have asked whether pre-verbal infants or nonhuman animals may be able to learn simple relationships in patterns of stimuli, and whether they can generalise these to novel sequences. Marcus *et al.* (1999) generated 3 simple AG structures (AAB, ABA and ABB, where both A and B elements were represented by a different randomly selected CV syllable, therefore ABA sequences may include 'le di le' or 'de je de', etc.). They showed that following habituation, 7-month old infants responded differently to AG structures different to that which they had been exposed to (e.g. following habituation to ABA, they responded more strongly to ABB or AAB), even when novel syllables were used. These results were initially viewed as evidence that infants developed 'algebraic rule learning' abilities earlier than traditionally believed (Marcus *et al.*, 1999). However, similar experiments in nonhuman animals suggest that a more parsimonious explanation may be that these abilities are not language specific at all. Murphy *et al.*, (2008) used Pavolvian conditioning to train rats (*Rattus norvegicus*) to respond only to certain sequences of stimuli (either patterns of illumination or tones of different pitches, following the structures AAB, ABB or ABA). They found that rats not only learned to respond only to the trained stimulus, demonstrating that they could discriminate between different sequence structures, but that these results generalised to novel stimuli (sounds of different pitches, Murphy *et al.*, 2008). Using similar sequences Hauser *et al.*, (2009) demonstrated that free ranging Rhesus macaques (*Macaca mulatta*) also show similar abilities. A wild population of monkeys was tested with sequences of the form AAB or ABB, where the A and B elements represented randomly sampled macaque vocalisations (i.e., the AAB structure could be presented as 'coo', 'coo', 'grunt' or 'bark', 'bark', 'pant-threat', etc.). Following habituation to one type of AG structure presented from a concealed audio speaker, Hauser and Glynn showed that the animals produced stronger dishabituation responses towards the other AG structure (i.e., those habituated to AAB responded more strongly to ABB, Hauser & Glynn, 2009). These studies suggest that human infants, rodents and nonhuman primates are able to learn these relatively simple, invariant patterns, and to generalise them to sequences of novel exemplars, suggesting some level of rule learning which could not be achieved through statistical learning alone.

Recognising and generalising patterns in sequences of stimuli is necessary for human syntax. These abilities, required for language but not unique to language processing, have been referred to as the 'language faculty in the broad sense', (Hauser *et al.*, 2002). One ability which has been claimed to be both necessary for, and unique to, human language, therefore falling into the 'language faculty in the narrow sense', is recursion (Hauser *et al.*, 2002; Fitch *et al.*, 2005; however also see Jackendoff & Pinker, 2005; Pinker & Jackendoff, 2005). A recursive structure is one in which a phrase can be embedded within another phrase, for example "the boy [the girl kicked] ran away". More formally, a recursive structure is a computational device which calls itself (Corballis, 2007). Being able to understand such recursive, hierarchical structures is argued to be fundamental to language, but does any equivalent ability exist in nonhuman animals? A number of studies have attempted to address this question using AGL paradigms. Fitch and Hauser (2004) developed two AG structures consisting of nonsense words (consonant-vowel syllables) split into two categories (A and B) based on the gender of the speaker (i.e. the A category consisted of syllables spoken by a female speaker, B category syllables were produced by a male speaker). They designed both a simple 'finite-state grammar' (of the form $(AB)^n$, which produces the sequence ABAB, where $n = 2$) and a more complex 'phrase-structure grammar', modelling simple recursion (A^nB^n, producing AABB sequences, in which they argue one AB phrase is embedded within another, A[AB]B). Fitch and Hauser showed that following exposure to the simpler FSG, cotton-top tamarin monkeys were able to discriminate between the two AG structures, responding more strongly to the novel A^nB^n

structure. However, following exposure to this more complex, 'recursive' AG structure, the monkeys showed no difference in responses when tested with the two AGs (Fitch & Hauser, 2004). These results led the authors to conclude that tamarins are capable of learning simpler structures, but since recursive abilities are not shared by nonhuman primates, the animals were unable to learn the more complex artificial grammar.

A similar experiment was conducted in European starlings (*Sturnus vulgaris*), a type of vocal learning songbird (Gentner *et al.*, 2006). Here an explicit learning task was used to train birds to discriminate between sequences constructed of starling song motifs generated by the $(AB)^n$ and A^nB^n structures. Gentner *et al.* demonstrated that starlings not only showed significant differences in responses to violations of both AG structures (up to n = 4, i.e. ABABABAB and AAAABBBB), but also that they were capable of generalising these to sequences comprising of novel song motifs (2006). Visual versions of the same AG structures have been tested in both pigeons (which are not vocal learners or songbirds) and kea (a species of parrot which does produce complex song Stobbe *et al.*, 2012). These experiments showed comparable results to those of Gentner *et al.* (2006), with both species recognising violations of both the $(AB)^n$ and A^nB^n structures (Stobbe *et al.*, 2012). Furthermore, kea, to a relatively higher level than pigeons, were able to generalise the rules governing these structures, for example to extended sequences or novel colours of stimuli. However, both these studies and that of Fitch and Hauser (2004) have faced severe criticisms. While it is possible to linguistically or mentally represent the A^nB^n structure as two centre-embedded AB phrases, A[AB]B, it is equally possible to simply solve the AG by counting the number of A and B elements, or by recognising the adjacent repetition of elements in the AABB but not the ABAB sequences (Perruchet & Rey, 2005; Corballis, 2007; van Heijningen *et al.*, 2009). These results are therefore insufficient to conclude anything about "recursive" sequence processing. In order to liken these structures to recursive aspects of natural language it is necessary to appreciate the relationship between the different A and B elements. In the sentence "the boy [the girl kicked] ran away", it is critical to understand the relationship between the different noun phrases and verbs presented. A listener needs to understand that the girl did the kicking and boy ran away, this is the very essence of syntax. The AG developed by Fitch and Hauser does not require this association between A and B elements. This issue has been addressed in more recent human studies, where specific A and B elements have been associated together based on acoustic or visual similarities, producing structures of the form $A_1A_2A_3B_3B_2B_1$ (Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009), but these more rigorous structures are yet to be tested in nonhuman animals (Perruchet & Rey, 2005). Indeed, a more recent study in zebra finch (*Taeniopygia guttata*) demonstrated that while these songbirds respond differently to sequences generated by $(AB)^n$ and A^nB^n structures, these responses could be explained in simpler terms (van Heijningen *et al.*, 2009). Specifically, the finches appeared to respond most strongly to individual pairs of elements (either the AA or BB bigrams, at either the beginning or end of the AABB sequences), regardless of whether these bigrams were part of AABB or AAAA sequences. Based on these results, van Heijningen *et al.*, concluded that the results from this study (2009), as well as those in tamarins and other birds (Fitch & Hauser, 2004; Gentner *et al.*, 2006; Stobbe *et al.*, 2012) fail to provide evidence for learning recursive structures.

All the studies discussed in this section require certain capabilities that are likely similar to abilities used in language (recognising word boundaries, identifying and generalising patterns of stimuli, categorising stimuli into A and B classes, etc.). However, once these processes have been performed, the sequences produced by the AGs are entirely invariant (i.e. the $(AB)^n$ structure will always produce the sequence ABAB, where $n = 2$). Human language is more complex than this, and even infants are not exposed to simple syntactic rules in isolation (Saffran, 2002), but must extract statistical regularities from sentences of varied length, composition and structure. A number of AG structures which aim to emulate this feature of language have been developed in human adults and infants (e.g., Reber, 1967; Saffran, 2002; Saffran *et al.*, 2008), and two studies have attempted to test this in nonhuman animals.

Saffran *et al.* (2008) presented both human infants and cotton-top tamarins with sequences generated by a forward-branching, nondeterministic AG structure. This AG contained 5 different nonsense word elements, which could occur in a range of constrained sequences, allowing the generation of a large amount of variation in sequence composition (see Chapter 2 and Chapter 3). Saffran *et al.* found that while human infants easily learned a number of variants of this structure, the tamarins only showed dishabituation responses to relatively simple versions of the experiment (Saffran *et al.*, 2008). Furthermore, in this experiment (but not those in human participants) the ‘grammatical’ test sequences, which were consistent with the AG structure, were the same as those used in the habituation phase, while the ‘ungrammatical’ violation sequences were obviously completely novel to the animals. Therefore it is not possible to determine whether novelty or a deeper understanding of the AG structure best explain the results. Bengalese finches have been tested using motifs from finch song following the same AG structure (Abe & Watanabe, 2011). Abe *et al.* reported that the finches were able to discriminate between sequences which follow the AG structure and those which violated it. Furthermore, using a modified version of the AG, they argued that the birds were able to learn a recursive, centre-embedded AG (Abe & Watanabe, 2011). However, these results have been challenged on the basis of serious acoustical confounds (the 'consistent' test sequences contained only minor variations relative to the habituation sequences, and were therefore acoustically very similar, while the 'violation' sequences had markedly different structures and were highly acoustically distinct, suggesting that the responses need not be attributed to any sort of recursive processing, Beckers *et al.*, 2011; Berwick *et al.*, 2012). These non-deterministic, branching artificial grammars have the potential to provide important insights into how nonhuman animals may process statistical regularities, and how these may relate to some of the abilities required to learn or process human syntax. However, potentially due to the difficulty of the task, it has proved challenging to rigorously demonstrate such AGL in nonhuman primates, and even vocal learning songbirds' responses may be based on relatively simple learning strategies or acoustical differences between conditions.

Artificial grammar learning paradigms provide a valuable opportunity to better understand the language related capabilities of non-linguistic species. The lack of reliance on other associated abilities, such as the association of semantic meanings to auditory or visual stimuli, allows the study of these abilities in a rigorous and controlled manner. Furthermore, shifting the emphasis of such experiments onto the perception rather than production abilities of the animals (Hurford, 2012; Petkov & Wilson, 2012) and the underlying computational mechanisms these rely upon (Hauser *et al.*, 2007) is necessary if we are to pursue such questions in species like nonhuman primates with very limited vocal repertoires (Petkov & Jarvis, 2012). However, a number of these studies have faced difficulties or challenges (Berwick *et al.*, 2011; Berwick *et al.*, 2012), leaving the abilities of nonhuman animals to learn artificial grammars unclear (Pena *et al.*, 2002; van Heijningen *et al.*, 2009). In Chapter 2, I will present an approach to quantify the complexity of these AG structures and to highlight the cognitive mechanisms or abilities minimally needed to identify consistent and violation sequences.

1.4. Neurobiology of language

Ever since Paul Broca's seminal work describing a region of the frontal cortex critical for speech comprehension a century and a half ago (Broca, 1861a), the neurobiological underpinnings of human linguistic abilities have been a subject of great interest. Broca believed that he had identified the seat of language in the brain, presenting evidence that individuals who had suffered lesions in the left inferior frontal gyrus (including, but not limited to, Brodmann Areas (BA) 44 and 45, Brodmann, 1909) showed dramatic language production deficits (Broca, 1861a; 1861b; 1861c; 1865; Dronkers *et al.*, 2007). He reported that these patients understood speech, and that their language abilities were still intact, however the vocabulary which they could produce was drastically reduced. This disorder became known as Broca's aphasia, and the region which he described as Broca's area (Dronkers *et al.*, 2007).

A decade later, Carl Wernicke described two patients with severe word comprehension disorders who produced paraphasic speech (they produced fluent speech which contained many unintended syllables, words or phrases, Wernicke, 1874; Eggert, 1977). Wernicke argued that these deficits were due to impaired auditory images of words; the patients were unable to associate the appropriate words with their internal representations. Autopsies revealed that these patients had suffered lesions in posterior parts of the left superior temporal gyrus, leading to the claim that this region may represent a lexicon for spoken words. This region and the disorder associated with damage to it are typically referred to as Wernicke's area and Wernicke's aphasia. This research highlighted an interesting dissociation with the discoveries of Broca. Broca's aphasia represents an inability to produce speech, while speech comprehension is relatively maintained, Wernicke's aphasia however is characterised by relatively normal speech, with dramatic impairments to vocal perception.

These 19th century lesion studies provided the initial foundations for our understanding of how language is processed in the brain and the idea that specific regions of the brain are linked to particular cognitive mechanisms and language related abilities (Broca, 1861a; Wernicke, 1874). While such studies may provide valuable insights into the neurobiology of the language network, lesions are rarely focal enough to provide a high degree of specificity about the involvement of precise regions, and plasticity and recovery make identifying normal brain function more difficult (Friederici, 2011). More than a century of research and technical developments (e.g. MRI, EEG, MEG) have drastically improved our understanding of language processing and the neurobiological network which supports it.

1.4.1. The language network

Human language is a uniquely complicated communication system, and the brain network supporting it is similarly complex. However, while many details remain uncertain, there is a high level of agreement about some aspects of the network (e.g., Damasio & Geschwind, 1984; Friederici, 2002; Vigneau *et al.*, 2006; Hickok & Poeppel, 2007; Friederici, 2011). Language production and processing primarily activate significant areas of the left perisylvian area, comprising of the IFG (including Broca's area), anterior, middle and posterior temporal lobe and temporo-parietal areas and a number of subcortical regions including the basal ganglia (Friederici, 2002; Vigneau *et al.*, 2006; Hickok & Poeppel, 2007; Tyler & Marslen-Wilson, 2008; Friederici, 2011). These regions are connected by at least two distinct pathways: a dorsal pathway between the posterior temporal lobe and Brodmann Area 44 in the IFG; and a ventral pathway between the anterior temporal lobe and BA45 and other ventral IFG regions (Hickok & Poeppel, 2004; Friederici *et al.*, 2006a; Hickok & Poeppel, 2007; Catani & Thiebaut de Schotten, 2008; Saur *et al.*, 2008; Friederici, 2009; Rauschecker & Scott, 2009; Saur *et al.*, 2010), which may in turn be comprised of additional separate pathways (e.g., see: Catani *et al.*, 2005; Friederici, 2009; Friederici, 2011; Petkov & Wilson, 2012). The network of brain areas that support language processing has been the subject of much debate (for reviews see Friederici, 2002; Vigneau *et al.*, 2006; Hickok & Poeppel, 2007). Rather than repeating these discussions, this section will summarise the key brain regions that appear to be involved in the processing of different aspects of language and the pathways and connections between them. I will then discuss evidence for potentially language related abilities in nonhuman animals, and how these two fields could be integrated.

Auditory cortex

Initial cortical processing of all auditory stimuli occurs in the auditory cortex (AC), located bilaterally on the superior temporal gyrus. Primary regions of the AC, located in Heschl's Gyrus are involved in processing simple acoustic properties of sounds (e.g. pitch, Patterson *et al.*, 2002), while outlying secondary and tertiary regions respond to more spectrally and temporally complex stimuli (Hall *et al.*, 2002). While this initial auditory processing is critical to language perception, AC appears to be involved in the processing of all acoustic stimuli, including speech sounds and pitch, with little specialisation for language above other auditory stimuli (Zatorre *et al.*, 1992), although there is evidence for some speech specialisation in the left hemisphere (Zatorre *et al.*, 2002). Nevertheless, more complex, language specific processing also recruits additional cortical regions.

Anterior temporal lobe

A number of studies have reported language-relate activation in the anterior temporal lobe (Friederici, 2002; Vigneau *et al.*, 2006; Hickok & Poeppel, 2007; Tyler & Marslen-Wilson, 2008; Friederici, 2011). This region appears to be particularly involved in syntactic processing at the phrase level, recognising the relationships between individual words or small phrases, rather than whole sentences (Mazoyer *et al.*, 1993; Friederici *et al.*, 2000; Humphries *et al.*, 2005; Humphries *et al.*, 2006). Studies in a variety of languages have shown that when fMRI activation is contrasted between lists of pseudowords (with no syntactic structure or semantic meaning) and pseudoword sentences (which also have no semantic meaning, but syntactic structure is preserved), activation is observed in the anterior temporal lobe (as well as posterior areas, temporo-parietal regions and the IFG, Mazoyer *et al.*, 1993; Vandenberghe *et al.*, 2002; Humphries *et al.*, 2005; Humphries *et al.*, 2006). Furthermore, when real words or sentences were compared to pseudowords or pseudoword sentences respectively, an increase in activation was observed in temporo-parietal regions and some areas of the temporal lobes, but not in anterior temporal regions, suggesting this area is not involved in semantic processing (Friederici *et al.*, 2000; Humphries *et al.*, 2006). Further evidence suggesting specialisation for syntactic, rather than semantic processing comes from observations that the anterior temporal lobe is activated when participants were presented with sentences containing syntactic violations but not semantic ones (Friederici *et al.*, 2003). Finally, a meta-analysis of 120 fMRI studies of semantic processing of individual words, highlighted a broad left lateralised network of brain regions, but showed no consistent activation in anterior temporal regions (Binder *et al.*, 2009).

While at the level of phrases or individual words, anterior temporal regions appear to be primarily involved in the processing of syntax, evidence suggests that they may play a role in processing semantics in other circumstances. Vandenberghe and colleagues presented participants with a variety of sentence stimuli (Vandenberghe *et al.*, 2002). They found that the anterior temporal lobe responded more strongly to sentences than scrambled sentences, reiterating the area's involvement with syntactic processing. However, they also found that when scrambled sentences were compared to similarly scrambled sentences in which some of the words were replaced with random words (therefore containing both syntactic and semantic violations) increased anterior temporal lobe activity was observed (Vandenberghe *et al.*, 2002). Furthermore, Rogalsky and Hickok investigated syntactic and semantic processing by asking participants to judge whether sentences contained either semantic or syntactic errors (Rogalsky & Hickok, 2009). In their analysis, only sentences containing no violations were considered, so any differences in activation could be explained only by attention on a particular aspect of the sequence. They found that significant regions of the anterior temporal lobe were involved in both tasks, but also identified a small sub-region of the anterior temporal lobe that was preferentially activated in the semantic condition. These results led the authors to conclude that the anterior temporal lobe is involved in the processing of both semantic and syntactic information at the sentence level (Rogalsky & Hickok, 2009). These studies suggest that at the phrase level the anterior temporal lobe seems to be involved in syntactic but not semantic processing, however, at the sentence level it may also have some semantic functions.

Posterior temporal lobe

The posterior parts of the temporal lobe were, after Broca's area, some of the first to be functionally linked to language comprehension and production (Wernicke, 1874). The posterior temporal lobe, as well as temporo-parietal regions and the angular gyrus, has since been implicated in a wide range of language related processes, including both syntactic and semantic processing (e.g., Friederici *et al.*, 2003; Kinno *et al.*, 2008; Friederici *et al.*, 2010). For example, Friederici *et al.* (2003) reported that sentences containing semantic violations activate a network of areas including the posterior and anterior temporal lobe, while syntactic violations produce activity in the left posterior temporal lobe, as well as the anterior temporal lobe, frontal opercular regions, and basal ganglia. The overlap in recruitment of the posterior temporal lobe in responses to both syntactic and semantic violations suggests that this area is not only involved in processing both aspects of language, but may also be involved in syntactic and semantic integration (Friederici *et al.*, 2003).

When presented with a range of sentences varying in terms of both syntactic complexity and memory demands, Cooke *et al.* (2002) report that the left posterior temporal lobe is the only region consistently recruited in the processing of all sentences, highlighting its central role in sentence comprehension. Furthermore, the posterior temporal lobe has been suggested to be activated by increasing sentences ambiguity (e.g. 'she weighs the flour' compared to 'she sifts the flour', Obleser & Kotz, 2010). Higher levels of ambiguity make predicting later words in the sentence more difficult, requiring more information to be held in memory to allow the syntactic and semantic integration (Rodd *et al.*, 2005; Snijders *et al.*, 2009; Obleser & Kotz, 2010). When participants were presented with relatively complex sentences with hierarchical structures, in contrast to simpler, more linear sentences, posterior temporal lobe activation is also observed (Bornkessel *et al.*, 2005; Friederici *et al.*, 2009).

These studies suggest that the posterior temporal lobe is involved in the processing of both syntactic and semantic features of language, but seems particularly involved in the integration of these two key aspects of language (Grodzinsky & Friederici, 2006; Friederici *et al.*, 2009; Friederici, 2011). However, while these imaging studies, and earlier lesion studies, clearly highlight the involvement of the posterior temporal lobe in language perception, other areas of the brain are typically recruited by similar tasks (Tyler & Marslen-Wilson, 2008; anterior temporal lobe, Obleser & Kotz, 2010; Snijders *et al.*, 2010; e.g., IFG, Friederici, 2011). Furthermore, this region's role in integration of features may not be limited to syntactic and semantic elements of language (Friederici, 2011), as the area has also been implicated in other forms of integration (e.g., audio-visual integration, Calvert, 2001; Amedi *et al.*, 2005).

Inferior frontal gyrus

The Inferior Frontal Gyrus (IFG), including Broca's area in the left hemisphere, is critical to language processing (Broca, 1861a). While there is general agreement that this region is involved in processing sentences of increasing linguistic and syntactic complexity (e.g., Friederici *et al.*, 2006b; Friederici, 2011), the specific function and even anatomical boundaries of Broca's areas, and sub-regions thereof, are debated (Hagoort, 2005; Grodzinsky & Santi, 2008; Rogalsky & Hickok, 2009; Amunts & Zilles, 2012). In this section I will outline attempts to delineate the anatomical and functional correlates of Broca's area and the inferior frontal gyrus more broadly, and discuss their involvement in language processing.

Broca's area (or territory) is typically anatomically defined as Brodmann areas BA44 and BA45 in the left inferior frontal gyrus (Brodmann, 1909; Friederici, 2011). However, when Broca's area is considered in functional, rather than anatomical terms (i.e., as the region of the IFG involved in language processing), its anatomical boundaries are more difficult to define (Amunts & Zilles, 2012). Beyond BA44 and BA45, the IFG consists of a number of anatomical areas which are thought to be involved in language processing (e.g. BA46, BA47, frontal operculum, insula, Friederici, 2004; Friederici *et al.*, 2006a). Even these anatomical areas may contain further subdivisions, for example into anterior and posterior parts of BA45 (BA45a and BA45p), or dorsal and ventral subdivisions of BA44 (44d and 44v, Friederici, 2011). Furthermore, anatomical boundaries vary between individuals, making precise correspondences between anatomical regions and brain functions difficult to map (Anwander *et al.*, 2007). Therefore, in this section, I will discuss the involvement of the IFG generally as well as evidence for the localisation of specific linguistic functions (aspects of language, such as syntactic movement) to individual regions, including Broca's area.

The IFG is involved in processing language at a number of levels. It is implicated in relatively low level processes, such as action observation and execution, (Pulvermuller & Fadiga, 2010), maybe homologous to the mirror neuron system observed in nonhuman primates (Rizzolatti & Arbib, 1998). However, while these functions may be critical to language acquisition, production or comprehension, these alone are clearly insufficient to explain the processing of the complex syntactic and semantic information contained in human language (Chomsky, 1957). At a higher level, the IFG is reported to be responsible for verbal Working Memory (WM) (Caplan & Waters, 1999; Wager & Smith, 2003; Rogalsky *et al.*, 2008). Verbal WM is critical to the processing of human language. Understanding the meaning of a sentence involves tracking the temporal positions of many words and phrases which must be syntactically combined to comprehend the intended meaning of a sentence (Just & Carpenter, 1992; Gibson & Pearlmutter, 1998; Cooke *et al.*, 2002; Lewis *et al.*, 2006). Furthermore, sentences of increasing complexity are likely to correspond with greater working memory demands, as longer sentences require more complex syntactic understanding (e.g., Santi & Grodzinsky, 2007). There is debate regarding whether the working memory system used in language processing is specific to syntax (Caplan & Waters, 1999; Fedorenko *et al.*, 2006; Santi & Grodzinsky, 2007) or reflects more general WM capacity (Waters & Caplan, 1996; Lewis *et al.*, 2006). However, there is a general consensus that language requires verbal WM, and that it is seated in the IFG. Some authors have suggested that the sole function of the IFG in language relates to working memory (Rogalsky & Hickok, 2009; 2011). However, a far greater number of imaging and patient studies suggest that, while the IFG is involved in WM, its function in language processing is more general (Caplan & Waters, 1999; Friederici, 2002; Opitz & Friederici, 2003; 2004; Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009). Furthermore, it has been proposed that WM and the processing of syntactic complexity may even have different anatomical correlates in the IFG, with more dorsal regions supporting memory and more ventral areas syntactic processing (Fiebach *et al.*, 2004; Makuuchi *et al.*, 2009). While there may be debate regarding whether the WM system in the IFG is unique to syntax or more general; and how this system interacts with complex syntactic processing, it is clear that both processes recruit IFG regions and both are required for language processing (Friederici, 2011).

There is a large amount of evidence that activation in this area increases with syntactic complexity. Syntactic complexity can be manipulated in a number of ways. For example, in languages which employ case-markings it is possible to create grammatically correct, non-canonical sentences that are more syntactically complex than their semantically equivalent, canonical counterparts. Linguistically, this is called scrambling, and has shown to increase activation in the IFG, including both BA44 and BA45 (Bornkessel *et al.*, 2005; Friederici *et al.*, 2006b; Kinno *et al.*, 2008). Alternatively, in languages that typically employ subject-first structure (e.g. in English, "the boy [*subject*] kicked [*verb*] the ball [*object*]") increased syntactical complexity is observed in non-canonical, object-first sentences, such as "the ball [*object*] was kicked [*verb*] by the boy [*subject*]". This operation is known as movement. A number of studies have shown increased IFG activation (BA44 and 45; as well as commonly showing temporal lobe activation) in responses to these non-canonical sentences relative to canonical ones (e.g., Ben-Shachar *et al.*, 2003; Ben-Shachar *et al.*, 2004; Constable *et al.*, 2004; Santi & Grodzinsky, 2007). Activation in the IFG has also been shown to correlate to the amount of movement in a sentence (i.e. more activation is observed in response to a longer antecedent–gap relation, Cooke *et al.*, 2002; Fiebach *et al.*, 2005). Higher levels of syntactic complexity are also observed in sentences containing centre-embedded phrases compared to those with typical, right-branching structures (Hauser *et al.*, 2002; Fitch & Hauser, 2004). For example "the boy [the girl kicked] ran away" is syntactically more complex than "[the girl kicked] [the boy who ran away]" (parentheses denote phrases). This is the same type of sentence structure that artificial grammars modelling 'recursive' sequences attempted to emulate (Fitch & Hauser, 2004; Gentner *et al.*, 2006; see Section 1.3 and below). A number of studies have shown that when the syntactically more complex, centre-embedded sentences are compared to their simpler counterparts, focal IFG clusters are observed, particularly in BA44 and BA45 (Just *et al.*, 1996; Stromswold *et al.*, 1996; Makuuchi *et al.*, 2009; Santi & Grodzinsky, 2010)

Artificial Grammar Learning in the IFG

Artificial grammar learning paradigms, similar to those discussed in nonhuman animals in Section 1.3.1, have also been used to assess the brain areas involved in processing sequences designed to emulate features of language (Petersson *et al.*, 2004; Forkstam *et al.*, 2006; Friederici *et al.*, 2006a). These studies have consistently demonstrated that violations of both auditory and visual artificial grammars produce activation in areas of the brain associated with language processing, particularly the IFG, suggesting that similar neural substrates might be involved in processing both natural language and these simpler structures (Gomez & Gerken, 2000; Friederici, 2002; 2004; Petersson *et al.*, 2004; Petersson *et al.*, 2012). Violations of artificial grammars that emulate more complex, non-adjacent or hierarchical relationships between words (Petersson *et al.*, 2004; Forkstam *et al.*, 2006; Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009; de Vries *et al.*, 2011) have been shown to recruit Broca's area (BA44 and BA45) during both explicit decision making (e.g. Petersson *et al.*, 2004; Forkstam *et al.*, 2006; Friederici *et al.*, 2006a) and implicit tasks (Folia *et al.*, 2011).

Beyond showing similar patterns of activation to features of natural language, the relatively simple nature of AG structures mean that participants are able to learn them quickly, presenting the opportunity to assess language learning in a way that is not possible with natural language. For example, Opitz and Friederici (2003; 2004) presented participants with a previously unfamiliar AG designed to emulate some key elements of natural language (Friederici *et al.*, 2002) during fMRI scanning. They demonstrated a dynamic pattern of brain activation, where initial acquisition of the AG structure appeared to strongly recruit hippocampal areas, while increased proficiency with the language, later in the scanning session, was associated with reduced hippocampal and increased IFG activation (Opitz & Friederici, 2003; 2004). These studies highlight the value of AGL paradigms to evaluate how the brain processes complex language related tasks in individual participants, demonstrating the difference between initial similarity-based strategies and ultimately more complex rule based learning.

The ease with which participants are able to learn artificial grammars relative to natural language has also been exploited by combining AGL paradigms with non-invasive brain stimulation techniques. Repetitive Transcranial Magnetic Stimulation (rTMS) and transcranial Direct Current Stimulation (tDCS) were used to stimulate BA44/45 during either artificial grammar learning (de Vries *et al.*, 2010) or testing (Udden *et al.*, 2008). In both experiments this stimulation produced enhanced abilities and faster reaction times in identifying artificial grammar violations, providing further evidence that Broca's area is causally involved in AGL.

A number of studies, motivated in part by some of the previously discussed work in nonhuman animals (e.g. Hauser et al., 2002; Fitch & Hauser, 2004; see Section 1.3.1), have used AGL paradigms to explore how the brain processes AG structures that aimed to emulate the centre-embedding and recursion seen in natural language (Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009). Friederici and colleagues used AGs similar to that of Fitch and Hauser (see Section 1.3, Fitch & Hauser, 2004). Participants were habituated to either a simple, relatively linear structure of the form $(AB)^n$, which produces sequences such as $A_1B_1A_2B_2$, or a more complex, hierarchical structure, A^nB^n, which produces sequences of the form $A_1A_2B_2B_1$, to assess how the brain may process these more complex AGs. It is important to note that the majority of these studies ensured the relationship between pairs of A and B elements were maintained (i.e., A_1 is associated with B_1) by testing participants with (violation) sequences including $A_1A_2B_1B_2$ (Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009), therefore the same criticisms directed towards some of the previously discussed animal work (e.g. Perruchet & Rey, 2005; van Heijningen *et al.*, 2009) are not applicable here. All of these studies showed that violations of both the relatively linear $(AB)^n$ structure and the more complex A^nB^n structure produced activation in ventral, frontal opercular regions of the IFG. However, only violations of the A^nB^n structure activated BA44 and BA45 (Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009). These results are comparable to studies using natural language sentences with similar structures (Just *et al.*, 1996; Stromswold *et al.*, 1996; Makuuchi *et al.*, 2009; Santi & Grodzinsky, 2010), which showed that in contrast to more linear structures that activated ventral IFG regions, centre-embedded sentences produced activation in Broca's area. Furthermore, when sentences were orthogonally varied to contain different levels of centre embedding or linguistic movement while maintaining working memory demands, dissociations between the two features were observed (Makuuchi *et al.*, 2009; Santi & Grodzinsky, 2010). FMRI activity associated with movement appeared primarily in BA45 while centre-embedding appears to recruit BA44.

These results demonstrate the involvement of ventral portions of the IFG in all AG tasks and also highlight the important role of Broca's area in processing complex AG structures. This has led to important hypotheses regarding the evolution of the language faculty in humans. AGL paradigms do not require linguistic stimuli, and comparable results have been observed with both words and visual symbols, unrelated to language (Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009). Therefore, it has been argued that some of the sequence processing abilities based in the IFG may not be specific to language (Friederici, 2002; 2004; Friederici *et al.*, 2006a; Petersson & Hagoort, 2012). These studies have highlighted that while ventral frontal cortex regions are involved in most AGL tasks, Broca's area is critically involved in processing of more complex AG structures, either involving centre-embedding and recursion or more complex and varied finite-state grammars. These results, in conjunction with natural language neuroimaging and lesion studies, suggest that Broca's area may be a critical region for processing more complex aspects of syntax. Furthermore, nonhuman primate AGL results have suggested that monkeys are able to learn simpler AG structures (Fitch & Hauser, 2004), but not more complex AGs that would activate Broca's area in humans (Friederici *et al.*, 2006a). This has led to the hypothesis that some of the human unique aspects of language may be supported by Broca's area, while simpler abilities require more ancestral ventral IFG regions, homologues of which are more likely to exist in nonhuman primates (Sanides, 1962; Friederici, 2004; Friederici *et al.*, 2006a; Friederici, 2011). However, evidence regarding the abilities of nonhuman animals to learn complex or centre-embedded AG structures is currently lacking (see Section 1.3.1, van Heijningen *et al.*, 2009; Berwick *et al.*, 2011; Beckers *et al.*, 2012; Hurford, 2012). Furthermore, no neuroimaging studies investigating the AGL capabilities of nonhuman primates have been conducted. Therefore, whether AGL in these animals might recruit presumed homologues of Broca's area (BA44/45), more ventral IFG regions, or some other neural substrates remains an empirical question.

1.4.2. Language pathways in the brain

As this chapter has discussed, language processing is highly complex and is supported by a number of interacting brain areas, the roles of which may be linked to specific features of language. The key regions in these processes are the inferior frontal gyrus (including both Broca's area in BA44 and BA45 and more ventral IFG regions), the anterior temporal lobe and posterior temporal and temporo-parietal regions (Wernicke's area). The first structural connection identified between Broca's area in the IFG and Wernicke's region in the posterior temporal lobe was the pathway known as the arcuate fasciculus (AF) in 1895 (Dejerine, 1895). Since the discovery of this dorsal pathway, neuroimaging techniques have also revealed ventral connections, via the Extreme Fibre Capsule System (EFCS), between the anterior IFG, including BA45 and BA47, and the anterior temporal lobe (e.g., Catani & Thiebaut de Schotten, 2008).

A number of studies have reported that the ventral stream, between anterior parts of the temporal lobe and ventral IFG regions, is involved in processing sound-to-meaning mapping; connecting auditory input with appropriate semantic labels (Hickok & Poeppel, 2004; 2007; Saur *et al.*, 2008; Rauschecker & Scott, 2009; Saur *et al.*, 2010). However, using probabilistic fibre-tracking approaches (Diffusion Tensor Imaging, DTI) Friederici and colleagues identified a potentially discrete ventral pathway, the Uncinate Fasciculus (UF), connecting the frontal operculum to the anterior temporal lobe (Friederici *et al.*, 2006a). The seed region for this analysis was based on a brain area that showed preferential activation to simple (adjacent) AG violations, therefore the authors concluded that this pathway may be required for processing local relationships between neighbouring elements in a sequence. Therefore, it appears that the anterior temporal lobe may be connected to different ventral inferior frontal regions by two different pathways, potentially with different functional roles (Friederici, 2011).

The dorsal pathway connects the posterior temporal lobe to the premotor cortex and inferior frontal gyrus via the arcuate fasciculus. This pathway has been implicated in the auditory motor integration, and the sensory-motor mapping of sound articulation (Hickok & Poeppel, 2004; 2007; Saur *et al.*, 2008; Rauschecker & Scott, 2009; Saur *et al.*, 2010). However, the same DTI approach that revealed the ventral, UF pathway in response to simple syntactic violations, highlighted a second dorsal stream, connecting BA44 in the IFG to posterior temporal regions, in response to more complex, non-adjacent AG violations (Friederici *et al.*, 2006a). This dorsal pathway (the superior longitudinal fasciculus, SLF) appears to originate (or terminate, DTI approaches are insufficient to determine directionality) in BA44, anterior to the premotor areas identified in previous studies, suggesting that the dorsal stream may also involve two distinct pathways (Friederici, 2011). These results are supported by the evidence that the first dorsal pathway (AF), between the posterior temporal lobe and premotor cortex, is present in infants (Dubois *et al.*, 2006), while the second dorsal pathway (SLF), involving BA44, develops much later (Brauer *et al.*, 2011). Strong connections between sensory information and motor output, which would likely require the AF, are necessary for initial language learning, so are unsurprisingly present in young infants. However, more complex syntactic abilities, which may be supported by the second dorsal stream (SLF), develop more slowly (Hickok & Poeppel, 2007; Dittmar *et al.*, 2008; Friederici, 2011).

The brain areas and connections between them, which are responsible for supporting language processing in the brain, are complex and currently imperfectly understood. However, it is clear that language relies upon the connected processing of a distributed perisylvian network, centred around the IFG connected by two sets of pathways to the anterior and posterior temporal lobe.

1.5. Language related neurobiology in nonhuman animals

The network of human brain areas involved in the processing language has been the subject of more than a century of research. While questions remain regarding the fine detail of the roles of these different regions, and how they interact to process different features of language, a broad consensus has emerged. However, far less research has considered how communication or language related abilities might be represented in the brains of nonhuman animals.

Songbirds have become a valuable model species in researching certain language related abilities. Due to their impressive vocal learning and song production capabilities (Jarvis, 2004; Petkov & Jarvis, 2012), both their natural song production and their AGL abilities have been the subject of investigation (Okanoya, 2004b; Abe & Watanabe, 2011). However, songbirds are very distantly related to humans, and their neurobiology is markedly different to that of primates, therefore any ‘language-like’ abilities observed in songbirds are likely the result of convergent evolution rather than common descent (Petkov & Jarvis, 2012). Therefore, while songbirds might show interesting behavioural similarities to humans, it is doubtful whether they have the potential to inform us regarding either the evolution of human language or its neuronal underpinnings.

Relative to humans and even other vocal learning species such as songbirds, nonhuman primates show very limited vocal ability. As has previously been discussed, nonhuman primates are at best limited vocal learners and do not show the same level of production abilities as is observed in some avian species (see Section 1.2.1). However, many monkey species do engage in vocal communication. Recent developments in neuroimaging technology (Ogawa *et al.*, 1992; Logothetis *et al.*, 1999; Van Meir *et al.*, 2005; Logothetis, 2008; Poirier *et al.*, 2009) have allowed several groups to reveal how vocal communication signals are processed in the brains of nonhuman primates (Poremba *et al.*, 2004; Gil-da-Costa *et al.*, 2006; Petkov *et al.*, 2008b; Taglialatela *et al.*, 2008; 2009). While natural language undoubtedly communicates more information in more complex ways than do the communication calls of nonhuman animals, a recent shift in focus has highlighted that some of the information communicated may be comparable across species. For example, a number of studies have investigated how properties of vocalisations (speech in humans or different calls in nonhuman primates) are processed in the brains of different species, with particular reference to the acoustical properties of the sounds (Dehaene-Lambertz *et al.*, 2005; Liebenthal *et al.*, 2005; Rimol *et al.*, 2005; Obleser *et al.*, 2006; Obleser *et al.*, 2007; Taglialatela *et al.*, 2009). Studies in humans, chimpanzees and Rhesus macaques, have demonstrated that hearing the vocalisations of conspecifics causes activation in comparable areas of the superior temporal lobes in both hemispheres (Wilson & Petkov, 2011). Furthermore, conspecific vocalisations relative to those of other species, or of known individuals compared to unknown conspecifics, both recruit an anterior temporal lobe 'voice' region in both humans (Belin *et al.*, 2000; Belin & Zatorre, 2003; von Kriegstein *et al.*, 2003) and macaques (Petkov *et al.*, 2008b) (unfortunately no data is currently available in chimpanzees). While certain features of human language might be unique, studies such as these have shown that some of the information that can be extracted from vocalisations appears to be comparable across different primate species.

Focussing comparative neuroimaging efforts on those basic, underlying properties of language, rather than the traditional perspective of concentrating on the unique features of language, allows direct comparisons to be drawn between species. This has a number of potentially important implications. These experiments provide novel insights into how the brains of nonhuman animals' process information contained within the vocalisations of conspecifics. Furthermore, comparative data of this type can inform us about the evolution of such abilities. If we observe that the same abilities caused activation in comparable areas of the brain, those results would suggest that such capabilities likely evolved by common descent, before the last common ancestor shared by the species. Cases in which both behavioural and neuronal correspondences exist between species represent the opportunity to study these abilities, in the nonhuman species, at a cell or molecular level that is not possible in human participants. For example, the identification of a voice sensitive area in a comparable region of the anterior temporal lobe in both humans (Belin *et al.*, 2000; Belin & Zatorre, 2003) and macaques (Petkov *et al.*, 2008b), has facilitated electrophysiological investigation of these areas in the macaque. Perrodin and colleagues, were able to target this region, which had previously been localised with fMRI, and discovered individual neurons that were specifically sensitive to voice information in monkey vocalisations (Perrodin *et al.*, 2011). This research represents a level of enquiry that is not possible in human participants, and highlights the exciting potential of taking a comparative approach to studying brain function.

1.6. Hypotheses and predictions regarding AGL in the nonhuman primate brain

Taking a comparative approach to neuroscience represents a genuine opportunity to investigate the evolutionary origins of a trait or ability, as well as the neurophysiological mechanisms that support it. The efficacy of this approach has been demonstrated by investigating how information represented in the vocalisations of different species is represented in the brain (see Petkov *et al.*, 2008b; Perrodin *et al.*, 2011). A similar approach, in combination with artificial grammar learning paradigms, may be able to inform us about the sequence-structure learning abilities of nonhuman animals, and how they relate to syntactic abilities in humans.

A large amount of research has demonstrated that AGL tasks recruit areas in the human IFG, consistent with some of those involved in processing natural language (e.g. Petersson *et al.*, 2004; Friederici *et al.*, 2006a; Petersson *et al.*, 2012). More complex artificial grammars, and increasing complexity in sentences, produce additional activation in Broca's area (BA44/45), while simpler structures appear to primarily recruit more ventral IFG areas including the Frontal Operculum (FOP). Artificial grammar learning experiments in nonhuman animals have consistently demonstrated that nonhuman primates, among other animals, are able to learn relatively simple AG structures (Fitch & Hauser, 2004; Newport *et al.*, 2004). More complex AGs, such as those reported to activate Broca's area in humans, have been insufficiently tested in nonhuman animals, but there is no conclusive evidence that nonhuman primates are able to learn such AG structures (van Heijningen *et al.*, 2009; Berwick *et al.*, 2012; Hurford, 2012). Based on this information, a number of predictions could be made regarding how the nonhuman primate brain may process artificial grammars.

The first possibility is that in response to violations of relatively simple AGs, nonhuman primates, like humans, would show increased activation in ventral IFG regions, homologous to the FOP in humans. It is possible that artificial grammars of sufficient complexity to activate Broca's area in humans are simply too complex to be learned by nonhuman primates. This might suggest that nonhuman primates do not possess a functional homologue of Broca's area, and therefore this region, and the capabilities that it supports, may be uniquely specialised for language in humans. Potential evidence for this hypothesis could be drawn from DTI data demonstrating less evidence for a dorsal pathway in chimpanzees and macaques than humans (Rilling *et al.*, 2008), however the existence of functional homologues of Broca's area in nonhuman animals currently remains an empirical question.

An alternative hypothesis is that AGs that would activate the FOP in humans might also activate homologues of Broca's area in nonhuman primates. Increasing syntactic complexity in both natural language (Makuuchi *et al.*, 2009; Santi & Grodzinsky, 2010) and artificial grammar structures (Friederici *et al.*, 2006a) is associated with increased activation in Broca's area in humans. It is possible that nonhuman primates may possess functionally comparable brain areas, but the threshold beyond which homologues of Broca's area may be recruited may be very different to that observed in humans. Quantifying the level of complexity of different artificial grammars represents a continuing challenge (Hurford, 2012; Jaeger & Rogers, 2012; Petkov & Wilson, 2012; and see Chapter 2), and comparisons across species are currently very limited. Therefore it is difficult to predict how different species, and their brains, might respond to different AG structures.

Finally, it is possible that completely different brain regions might be involved in AGL tasks in humans and monkeys. If the network of brain areas involved in language processing in humans represents a specific and unique evolution for language, then the behavioural AGL results observed in nonhuman primates might be supported by entirely different brain areas. This need not apply only to different regions of the IFG, but to all of the areas involved in language processing in humans including, particularly the anterior and posterior temporal lobes.

Comparative neuroimaging has the potential to reveal which of these hypotheses, if any, best reflect the brain areas involved in AGL in nonhuman primates. Empirical data is required to clarify whether brain regions homologous to those involved in human language are present in nonhuman species, the implications of these results on the evolution of the language faculty, and the potential for developing animal model systems in which the cellular mechanisms supporting these abilities can be explored.

1.7. Conclusions

In this chapter I have discussed the existing research and literature regarding how the AGL capabilities of nonhuman animals might inform us about language related abilities in humans. I have reviewed how language and artificial grammars are processed in the human brain. Finally I have discussed how these two fields could be integrated to investigate AGL in the nonhuman primate brain, and how this might ultimately inform us about the evolution of language and the neuronal mechanisms that support it. The aim of this thesis is to investigate the extent of the AGL capabilities of nonhuman primate species, and to comparatively assess how these are represented in the brains of both human and nonhuman primates. Better understanding of these areas may be critical to our understanding of the evolutionary origins of the language faculty, and in developing potential animal model systems in which such abilities can be investigated at a neuronal level. In Chapter 2, I will discuss some of the challenges involved with the comparative assessment of AGL abilities in a wide range of species, and present a novel, quantitative framework in which different AGs can be objectively compared. In Chapter 3, I will describe two experiments that show at both the group and individual level, that Rhesus macaques are able to learn a moderately complex, forward-branching, non-deterministic artificial grammar, and that these results cannot be easily explained by reliance of simple strategies. Chapter 4 compares the abilities of Rhesus macaques to more distantly related common marmosets and to human participants, suggesting a potential evolutionary gradient of syntactic complexity. Finally, in Chapter 5, I will describe fMRI experiments, in both Rhesus macaques and humans, demonstrating patterns of brain activation that are comparable, both between the species and to those observed during natural language processing in humans. I will conclude the thesis in Chapter 6, by discussing the implications of this research in terms of our understanding of the sequence-structure learning abilities of nonhuman primates, how this can inform us about the evolution of language, and the exciting implications of such research for developing animal models to study aspects of language at a cell and molecular level.

Chapter 2: Conceptual Challenges Relating to Comparative AGL

Parts of this chapter have been published as:

C. I. Petkov & **B. Wilson** (2012). On the pursuit of the brain network for proto-syntactic learning in nonhuman primates: Conceptual issues and neurobiological hypotheses. *Philosophical Transactions of the Royal Society B,* **367**, 2077-2088,

and:

B. Wilson, H. Slater, Y. Kikuchi, A. E. Milne, W. D. Marslen-Wilson, K. Smith & C. I. Petkov (2012). Auditory artificial grammar learning in macaque and marmoset monkeys. *The Journal of Neuroscience,* **33(48)**, 18825-18835.

All work presented in this chapter was primarily conducted by the author, except where explicitly stated.

Understanding the evolution of language presents a very difficult challenge (Christiansen & Kirby, 2003b). However, as discussed in Chapter 1, a comparative approach, including neuroimaging and AGL paradigms, may present an opportunity to clarify how the cognitive abilities of nonhuman primates might relate to the language network observed in the human brain. Previous research on this topic has been limited by a focus on the unique nature of human language and the obvious inabilities of nonhuman animals to process syntactic or semantic relationships in comparable ways to humans. However, recent developments in linguistics and modern language theory have led to advances in our understanding of language related processes (Bickerton & Szathmary, 2009; Tallerman, 2011; Fitch & Friederici, 2012; Hurford, 2012; Jaeger & Rogers, 2012), providing the opportunity to explore the features of language that nonhuman animals may be able to process.

In this chapter, I will discuss some of the conceptual challenges that have been faced in pursuing evolutionary precursors to human syntax in nonhuman animals. I will present a theoretically motivated approach to objectively compare different artificial grammars with relation to linguistic theory, so that the abilities of nonhuman animals can better be evaluated.

2.1. An evolutionary gradient of syntactic complexity

The Formal Language Hierarchy (FLH, Chomsky, 1957) or extended Chomsky hierarchy (Berwick *et al.*, 2011; Jaeger & Rogers, 2012) contains several categories of grammar, each describing an increasingly powerful computational language (Figure 2.1). The lower ranked grammars generate sets of languages that are subsets of those generated by higher ranked grammars. Therefore, the Finite State Grammars (FSGs) comprising finite and regular languages (1 and 2 in Figure 2.1) form a subset of the more complex Context Free Grammars (CFG; 3 in Figure 2.1) which in turn comprise a subset of the higher grammars. The most complex aspects of human language can be described context-sensitive or mildly context-sensitive grammars (4 and 5 in Figure 2.1, Berwick *et al.*, 2011; Hurford, 2012). Recursively enumerable languages represent a level of complexity only observed by computers or Turing machines (Chomsky, 1957). Nonhuman animals have been shown to learn simple FSGs (e.g. $(AB)^n$, Fitch & Hauser, 2004; Gentner *et al.*, 2006; van Heijningen *et al.*, 2009; Stobbe *et al.*, 2012). However, while numerous studies have claimed that various species are able to learn more complex CFGs (e.g. A^nB^n, tamarin monkeys: Fitch & Hauser, 2004; starlings: Gentner *et al.*, 2006; Bengalese Finches: Abe &

Watanabe, 2011) these results have been widely criticised for insufficiently demonstrating that the animals did not use simple learning strategies, (see Section 1.3.1, Perruchet & Rey, 2005; van Heijningen *et al.*, 2009; Berwick *et al.*, 2011; Beckers *et al.*, 2012; ten Cate & Okanoya, 2012). This focus on the ability of nonhuman animals to breach the potentially 'human unique' realms of CFGs (Hauser *et al.*, 2002), may have detracted focus from the range and variety of FSGs that nonhuman species may more easily be able to learn (Hurford, 2012). It is possible that a convincing demonstration of nonhuman CFG learning may arise (Jaeger & Rogers, 2012). However, it remains important to understand how the human capabilities to learn and produce CFGs and beyond may have evolved from abilities represented lower in the hierarchy, which may be present in extant nonhuman animals (Hurford, 2012; Jaeger & Rogers, 2012; Petkov & Wilson, 2012).

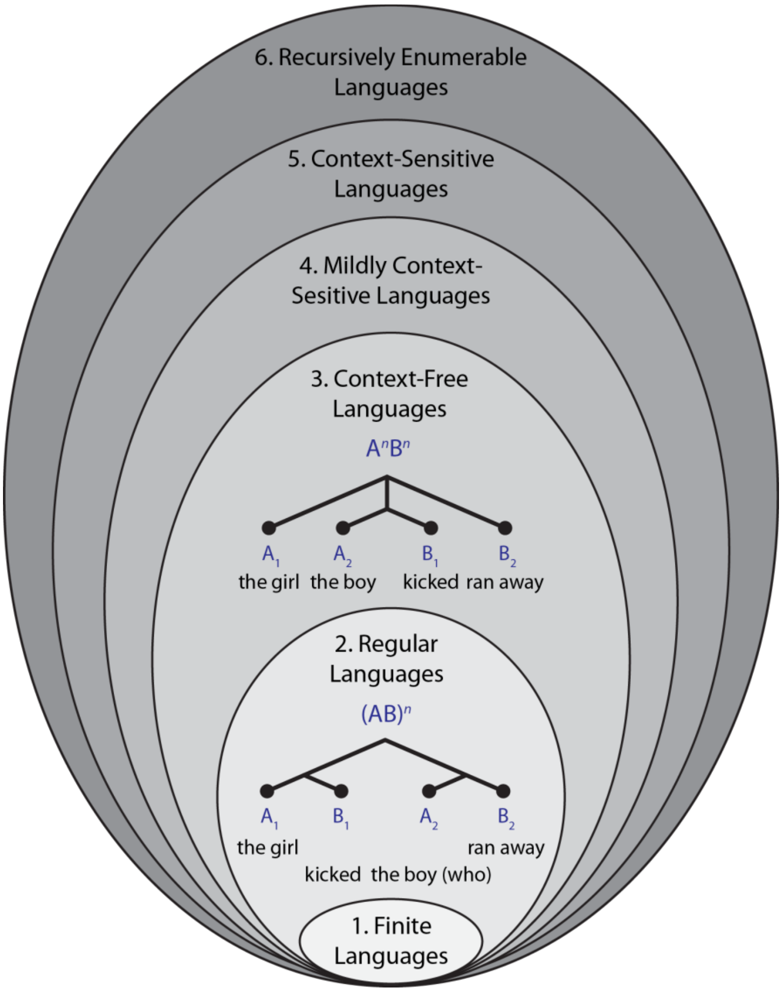


Figure 2.1 Chomsky hierarchy of grammars. Schematic of the Formal Language Hierarchy with examples of two key structures, (based on, Berwick et al., 2011; Petkov & Wilson, 2012). All animal species currently tested have been found to learn AGs equivalent in complexity to Regular Languages (2), while there is currently no strong evidence that any species other than humans can learn structures which can be modelled as Context Free Languages (3). Humans are able to process Mildly Context Sensitive (4) or possibly Context-Sensitive Languages (5) while Recursively Enumerable Languages (6) appear to be beyond the abilities even of humans and are only processed by computers or Turing machines.

2.1.1. Classifying vocal production abilities in nonhuman animals

How might the ability to generate context-free languages, or beyond, have evolved? One possibility is that when the ancestors to living humans began to organise vocalisations, and then words, into sentences of increasing complexity, this built upon the evolutionarily conserved ability to process sets of serially ordered strings. Then, at some point, selective pressures to reduce memory demands may have expanded syntactic capabilities by the adoption of rule-based learning strategies that avoid having to memorise all the elements and transitions in the sequences from more complex grammars (Teal & Taylor, 2000).

One of the simplest possible capacities is the ability to recognise and/or generate single elements. Such is the case for animals with call-based systems that can produce and recognise single vocalisations from a limited set of vocalisations (Figure 2.2). The next level of sequencing complexity is introduced when two calls are combined. In this case it becomes important to evaluate the relationships between adjacent elements. A subsequent level of complexity occurs when several elements are serially sequenced in a purely linear fashion. For example, zebra finches produce a linear combination of song motifs (Okanoya, 2004b). In such cases the pairwise transitions can be modelled by first-order Markov processes; any element can be predicted by the element immediately preceding it (Hurford, 2012) (Figure 2.2). Adding more elements or transitions to such a structure does not change the computational complexity of the pairwise sequencing process, but requires a larger indexical memory store.

At a higher level of complexity, some animals, such as Bengalese finches, nightingales and chaffinches, and humpback whales, produce songs that show sequencing elaborations such as forward- or backward-branching relationships and repeating elements (Hurford, 2012). These structures show more variation between species and songs than simpler structures, thus deviating from strictly linear processes (Honda & Okanoya, 1999). Rather than following an entirely predictable sequence, these structures allow variations where certain elements may be followed by two or more different elements. While representing a higher level of complexity, such structures still only require first-order Markov processes to model them (Figure 2.2). However, certain branching transitions, can be modelled either as a number of adjacent relationships, or could include more complex 'non-adjacent relationships' where an optional element can occur between two other elements. For example, in the forward-branching system in Figure 2.2, rather than learning all possible transitions between the A, B and C elements it may be simpler to learn that A will always be followed by C, but sometimes they will be separated by B. The recognition of non-adjacent relationships can reduce the need to memorise many pairwise transitions if the non-adjacency ‘rule’ can be learned. For adult humans, non-adjacent relationships can include even greater levels of complexity (e.g. nested or crossed relationships, Folia *et al.*, 2011; Udden *et al.*, 2011). Moreover, the ability to deal with non-adjacent relationships is not present at birth but seems to occur during infant development (Friederici *et al.*, 2011; Perani *et al.*, 2011). A final example of FSG complexity is the special case known as a state chain system (Hurford, 2012), in which the same element can occur at multiple positions within a sequence. Therefore, predicting the next element in the sequence requires knowledge of the preceding two elements (i.e., to identify which element will follow A in the state chain system in Figure 2.2, it is necessary to also consider the element that preceded A). Such transitions require higher order Markov processes, although much of the rest of the sequence could remain a first-order Markov process.

These examples help to illustrate the great variety seen in the song production of different animals’ and the variations in complexity of different ‘grammatical’ structures. However, an objective method to quantify the structural complexity of different artificial grammars would facilitate clearer comparisons between the abilities of nonhuman animals on different AGL tasks. If the learning abilities of different animals could be evaluated along quantitative dimensions of ‘syntactic complexity’, our understanding of the evolutionary bases for human syntactic abilities might be improved. In the following section, I will extend these ideas and propose a method by which the complexity of artificial grammars that have been used to test nonhuman animals could be quantified, and therefore how the abilities of nonhuman animals could be objectively compared.

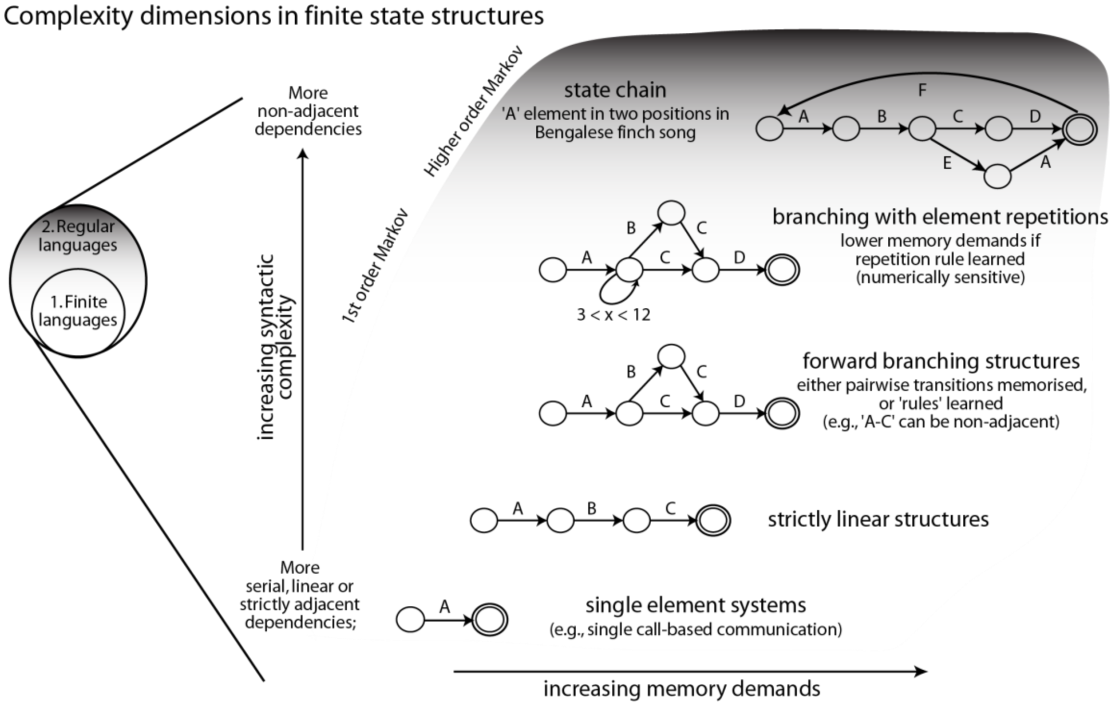


Figure 2.2 A conceptual framework for exploring the complexity of different structures. Adapted from Petkov and Wilson (2012). Schematised for quantifying the different dimensions of syntactic complexity, on the vertical axis, a measure of linearity can be used as a function of increasing memory demands on the horizontal axis. At the lowest level are single element structures. These are followed by strictly linear structures, with multiple elements. A higher level of complexity includes forward branching structures with non-adjacent relationships, with or without element repetitions. These are followed by 'state chains' that cannot be fully learned with only first-order Markov processes; here the transition following each 'A' element depends on the preceding element.

2.2. A parameter space for comparing artificial finite state grammars

Like the vocal communication systems of different species, artificial grammars vary widely in complexity. To better understand and compare data collected from different AGL experiments, it would be beneficial to have a method to objectively compare these structures. In this section I will consider the dimensions on which artificial grammars vary and describe a quantitative parameter space in which different finite state AGs can be objectively compared.

One important variation in complexity between AGs is the number of different stimuli which contribute to the structure; this could be considered to be the size of the vocabulary of the artificial grammar. An AG consisting of only two different stimuli (for example a light being either on or off, Murphy *et al.*, 2008) is, at least on this dimension, less complex than a structure containing 5 different stimuli (e.g. different nonsense words, Reber, 1967; Saffran *et al.*, 1996b; Saffran *et al.*, 2008). Accordingly, the first dimension of the artificial grammar parameter space (Figure 2.3) is the number of different elements which contribute to the AG structure.

It is important to note that some artificial grammars are based on a number of classes of stimulus, where the AG structure consists of only, for example ‘A’s and ‘B’s, but where ‘A’ and ‘B’ are themselves classes with several members. Therefore, learning the grammar initially requires the stimuli to be categorised into these two stimulus classes, before the structure of the AG (e.g. A^nB^n or $(AB)^n$, Fitch & Hauser, 2004) can be learned. This requirement for categorisation adds an additional level of complexity to learning an artificial grammar, and therefore must be considered in a complete model of the complexity of artificial grammars. Therefore, AGs requiring categorisation are represented as filled circles in Figure 2.3, while structures without categorisation are shown as open circles. Note that it could be argued that rather than being a binary distinction between whether an AG requires categorisation or not, it may be more appropriate to consider the number of stimuli which contribute to each stimulus class as an additional dimension for the parameter space. However, studies testing nonhuman animals have typically defined stimulus classes based on acoustic differences, for example the gender of the speaker producing nonsense words (Fitch & Hauser, 2004), or different song motifs in studies of songbirds (Gentner *et al.*, 2006). Therefore, the animals (unlike participants in some human artificial grammar learning experiments, Saffran *et al.*, 2008) are not required to individually learn which class a stimulus may belong to, only the rules governing stimulus class membership. Therefore, adding additional stimuli would be unlikely to increase the difficulty of the task, suggesting a binary classification is more appropriate.

A further key source of variation between AGs is the degree of predictability or determinism of the structure, reflecting the extent to which each element can be predicted by the preceding element(s). The sequence of elements (words or phrases) in human language is generally non-deterministic, making it important to understand how far nonhuman animals are sensitive to similar properties in the sequences generated by a given AG. As highlighted in the previous section, the songs of some songbird species can range from stereotyped and deterministic to much more variable, and this can be quantified by calculating their structural linearity (Honda & Okanoya, 1999). The same approach can be applied to artificial grammar structures. Linearity can be calculated as:

$$Linearity = \frac{Number\ of\ unique\ stimulus\ classes\ in\ the\ structure + 1}{Number\ of\ legal\ transitions\ to\ or\ from\ stimulus\ classes}$$

Linearity of 1.0 describes an entirely predictable, deterministic AG, where each category can be preceded and followed by only one legal transition. Lower values of linearity represent increased variability or unpredictability within a grammar, typical of more complex branching structures. The equation above includes transitions between stimulus classes and also to and from the start or end of the sequence. The number of stimulus classes considered in this equation contains an additional token representing the transitions from/to the Start or End of the sequence, so that a manifestly linear AG, e.g. Start → A → B → End, has 2 stimulus classes (A and B) and three transitions (→), where linearity is: (2+1)/3 = 1.0. The original Reber study in humans (Reber, 1967) contains 5 unique elements and 17 legal transitions, resulting in a linearity index of 0.35 (i.e., (5+1)/17), compared to 0.75 for $(AB)^n$ structures (i.e., (2+1)/4) (Fitch & Hauser, 2004). Therefore, the second dimension of the parameter space is the linearity of the artificial grammar (Figure 2.3).

The positions of artificial grammar structures which have been used to test nonhuman animals (Fitch & Hauser, 2004; Gentner *et al.*, 2006; Murphy *et al.*, 2008; Saffran *et al.*, 2008; Hauser & Glynn, 2009; van Heijningen *et al.*, 2009; Abe & Watanabe, 2011; Stobbe *et al.*, 2012), as well as, for reference, the original AGL study in humans (Reber, 1967), were mapped onto this parameter space. The abilities of animals to learn the different AGs are highlighted by checkmarks. In the following section I will discuss the abilities of different nonhuman species to learn different AGs, and how this can inform us about the cognitive mechanisms which underpin AGL and may represent some proto-syntactic abilities in nonhuman primates.

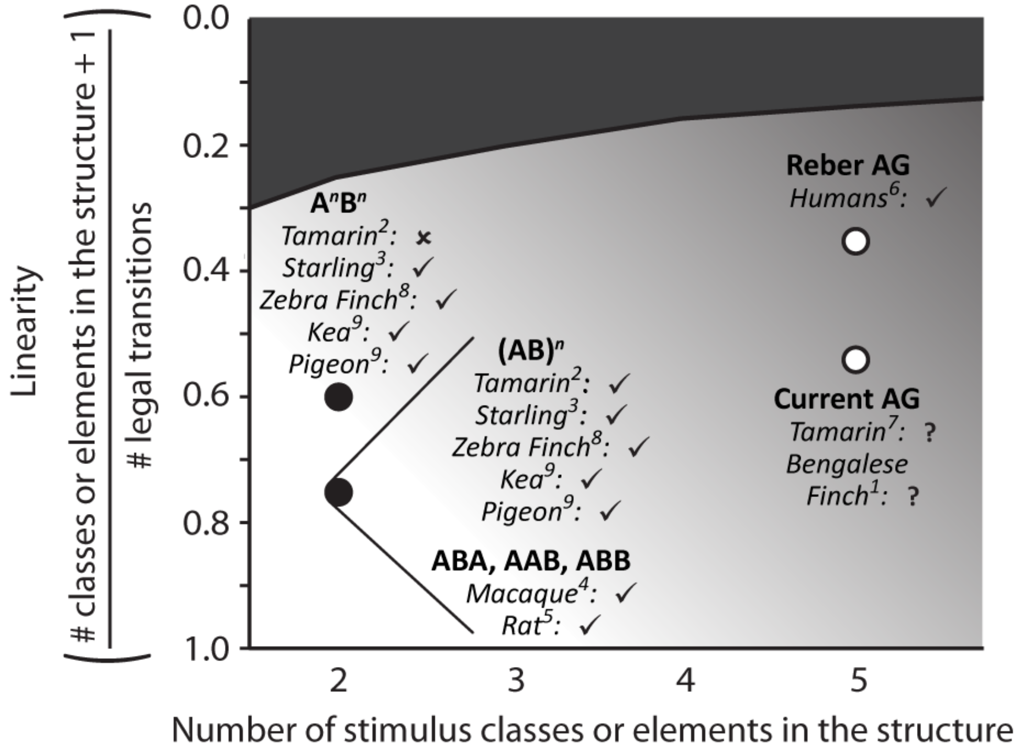


Figure 2.3 A parameter space mapping AG abilities in nonhuman animals. Mapping of AGs previously used to test nonhuman animals, including the original AG designed by Reber (Reber, 1967). These are plotted as the number of unique stimulus classes that contribute to the structure as a function of the linearity of the structure. The black line subdividing the shaded regions denotes the maximum possible structural non-linearity (i.e., random patterns devoid of structure). The checkmarks highlight AG structures which nonhuman species have been shown to learn. Crosses or question marks highlight uncertainty regarding whether the labelled species can learn those aspects. Figure references: 1: Abe & Watanabe, 2011; 2: Fitch & Hauser, 2004; 3: Gentner *et al.*, 2006; 4: Hauser & Glynn, 2009; 5: Murphy *et al.*, 2008; 6: Reber, 1967; 7: Saffran *et al.*, 2008; 8: van Heijningen *et al.*, 2009; 9: Stobbe *et al.*, 2012.

2.2.1. Mapping artificial grammars with two stimulus classes

Initially, AG structures containing just two stimulus classes, which have previously been used to test nonhuman animals, were mapped onto this parameter space. The AG structures $(AB)^n$ and A^nB^n, which produce the sequences, ABAB and AABB respectively (where n = 2, Fitch & Hauser, 2004; Gentner *et al.*, 2006; van Heijningen *et al.*, 2009; Stobbe *et al.*, 2012) and three-element long structures based on A/B classes of stimuli, (i.e., ABA, AAB and ABB, Murphy *et al.*, 2008; Hauser & Glynn, 2009), all contain only two stimulus classes, and therefore occupy the left side of Figure 2.3. The $(AB)^n$ and A/B structures are relatively linear, with only one transition which is not entirely predictable based on the prior element (e.g. $(AB)^n$ must begin with A, A is then always followed by B, and B can be followed by either A or 'End'), placing them towards the bottom left of the parameter space. As discussed in Section 1.3.1, there is evidence that many animals are able to implicitly or explicitly learn the $(AB)^n$ and the A/B type structures and may even generalise this learning to novel stimuli. However, in most of these studies it is possible that rather than learn the full extent of these AG structures, the animals were able to discriminate between consistent and violation sequences based on simpler learning strategies (e.g. Perruchet & Rey, 2005; van Heijningen *et al.*, 2009). Specifically, the A/B style AGL paradigms all involved testing individual animals with only two types of testing sequences at a time, (i.e., the structure they were exposed to or trained with, e.g. ABB, and one of the other two possible structures, ABA or AAB, Murphy *et al.*, 2008; Hauser & Glynn, 2009). If testing involves the ABA structure, it can always be differentiated from the other sequence by the absence of a repeated element. Even when tested with AAB and ABB sequences, a participant could simply identify whether the first two elements were the same and ignore the final element completely. Similarly, the tamarins were always tested with sequences of the form ABAB and AABB (Fitch & Hauser, 2004). Like the previous AGs, these can be solved by attending to only the first two elements in each sequence. It is important to note that more recent studies in birds used a wider array of test sequences and therefore overcame this problem (Gentner *et al.*, 2006; van Heijningen *et al.*, 2009; Stobbe *et al.*, 2012). However, even while taking a conservative view of the abilities of each species, it seems clear that all the animals tested are capable of categorising specific stimuli into A and B classes based on their acoustic properties (denoted by filled circles in Figure 2.3). Furthermore, while the mammalian species may have used relatively simple learning strategies, they are all at least able to detect whether pairs of elements follow a previously learned pattern. Therefore, the AGL capabilities mapped in the lower left of the parameter space appear to be common to all animals tested, and unsurprisingly are unlikely to be unique to human language.

The A^nB^n structure is less linear than $(AB)^n$, because while the sequence must always begin with A, A can be followed by either A or B, and B can precede either B or End, therefore it occupies a higher position in the parameter space (see Figure 2.3). (It is important to note that no AGL study in nonhuman animals has been sufficient to demonstrate more complex, context free grammar learning, Perruchet & Rey, 2005; van Heijningen *et al.*, 2009; see Section 1.3.1. Furthermore, as discussed below, this parameter space is designed to model only the adjacent relationships typically observed in finite state grammars, so is not necessarily appropriate to represent some of the more complex abilities demonstrated in humans, e.g. Bahlmann *et al.*, 2008) Experiments using this AG have produced more variable results in nonhuman animals, likely as a result of the accompanying increases in sequence learning demands. A number of avian species are able to recognise violations of these more non-linear structures through explicit training (Gentner *et al.*, 2006; van Heijningen *et al.*, 2009; Stobbe *et al.*, 2012), but cotton-top tamarin monkeys appear to be unable to implicitly learn the A^nB^n structure (Fitch & Hauser, 2004). These results suggest that tamarin monkeys find processing the A^nB^n structure difficult relative to the $(AB)^n$ structure, possibly as a result of increased nonlinearity. However, it is unclear whether differences observed between monkeys and songbirds in their abilities to learn this AG structure may result from the difference between learning by training compared to habituation (i.e., explicit vs. implicit forms of learning) or reflect a genuine cross-species difference. Studies that train animals to learn AG structures aim to tap into the reward-dependent learning system. Those using habituation/dishabituation paradigms and natural behavioural measurements aim to evaluate more implicit learning capabilities that need not engage perceptual awareness for learning to have occurred (Jarvis, 2004; Hurford, 2012; Petkov & Wilson, 2012). While these differences make comparisons between species more difficult, different methods of testing may be more appropriate than other in certain species. Explicit training of primates, particularly on auditory tasks, is a very difficult process, while measuring dishabituation responses is impractical in many small animals such as rodents and songbirds. It is therefore necessary to either compare results across different testing methods, or to attempt to test a species with a methodology to which it may not be well suited (even though other species are). Therefore, while it is possible to objectively compare the complexity of artificial grammar structures, it is also necessary to consider that inter-species differences observed may, at least in part, represent differences in testing procedures.

2.2.2. Mapping forward-branching, non-deterministic artificial grammars

The other AG structures, mapped in the right half of Figure 2.3, involve several elements or stimulus classes that contribute to the structure of the AG, offering considerable variation in the sequences that can be generated. As the number of elements or stimulus classes which contributes to an AG increases, so does the number of potential sequences which can be created. For example, the two stimulus class A^nB^n structure always generates the fixed sequence AABB (where $n = 2$). AG structures comprising of more elements can produce a variety of sequences and sequence lengths, e.g., 'TPTXVS' or 'VXVPXXVS' (Reber, 1967).

While AGs such as those inhabiting the upper right quadrant in Fig. 1A are learned with relative ease by human participants (Reber, 1967; Friederici *et al.*, 2002; Petersson *et al.*, 2012), they have not been tested with nonhuman animals, and their complexity may make them too difficult for them to learn. However, both cotton-top tamarins (Saffran *et al.*, 2008) and Bengalese finches (Abe & Watanabe, 2011) have been tested using simpler artificial grammars with multiple elements and a forward-branching structure. This AG shares many properties with the more complex structures used to test humans, but contains slightly fewer branches and is slightly more constrained, producing a more linear structure (although still less linear than the two stimulus class structures, Figure 2.3). The artificial grammar can be represented as:

$$Sentence = A + (D) + C + (G) + F + (C + (G))$$

Where A, C, D, F and G represent different nonsense-word elements, and parentheses denote optional elements (see Figure 3.1). In the first study, tamarin monkeys were habituated to sequences generated by the AG structure. They were then tested with consistent, legal sequences, and violation sequences that the grammar could not have generated. However, the only evidence for significant dishabituation responses to violations of the AG structure was obtained when the animals were tested with the same 'consistent' sequences to which they had been habituated (Saffran *et al.*, 2008). Thus the dishabituation responses of these New World monkeys may be based primarily on the unfamiliarity of the violation sequences.

Bengalese finches were tested using this AG as well as a slightly modified version of the structure (Abe & Watanabe, 2011). While the birds appeared to recognise violations of these AGs, several authors have noted that the testing sequences differed significantly in their acoustic properties between conditions. All consistent sequences were acoustically very similar to each other but the violation sequences differed considerably (Beckers *et al.*, 2012; Berwick *et al.*, 2012). Thus, the animals could have responded differentially to the test sequences based solely on acoustic differences. Based on these results it is difficult to assess the capabilities of nonhuman animals to learn such AG structures.

These experimental design issues make it very difficult to draw firm conclusions about the abilities of nonhuman animals to learn these more variable, non-linear AG structures. It is possible that both primates and songbirds are able to learn such grammars. However, more rigorous testing is required to support such conclusions and to inform us about how the abilities of different nonhuman animals might compare.

2.2.3. Mapping triplet segmentation studies

AGL experiments that investigate the abilities of human or nonhuman participants to identify word boundaries based on statistical relationships between syllables (e.g. Saffran *et al.*, 1996b; Aslin *et al.*, 1998; Hauser *et al.*, 2001) represent an important challenge to this model of AG complexity. As discussed in the previous chapter (see Section 1.3.1), these studies involve initially presenting participants with a stream of tri-syllabic nonsense words in which prosodic cues to word boundaries are minimised. Therefore successful identification of word boundaries relies on recognising that the statistical relationships between syllables within a “word” are stronger than those across word boundaries. Specifically, both human (Saffran *et al.*, 1996b; Aslin *et al.*, 1998) and primate (Hauser *et al.*, 2001) participants were presented with a continuous stream of synthesised consonant-vowel (CV) syllables arranged into tri-syllabic ‘triplet words’ (e.g. *tupiro, golabu, bidaku, padoti*). Each syllable only occurred in one word and only in one position, so ‘*tu*’ would always be followed by ‘*pi*’ while ‘*ro*’ could be followed by the initial syllable of any word. Hauser *et al.*, (2001) showed that following habituation to the stream of triplet words, cotton-top tamarins responded differently to ‘words’ that occurred in the stream compared to ‘part words’, which crossed word boundaries (e.g. *pirogo*) or ‘non-words’, which never appeared in the stream (e.g. *pikudo*). These results suggest that tamarins are not only sensitive to the order of the syllables in a synthesised speech stream, but that they are also able to derive statistical information from which they are able to detect ‘word’ boundaries. However, in order to plot the position of this AGL paradigm on a parameter space such as Figure 2.3, it is necessary to consider the animals' mental representations of the AG. The speech stream used by Hauser *et al.*, contained 12 unique syllables (Hauser *et al.*, 2001). One possibility is that the animals might have memorised 12 unique stimuli and the pairwise relationships between each of these, effectively processing an AG with 12 unique elements and a highly non-linear structure (Figure 2.4). However, evidence from songbirds suggests that rather than perceive each song motif independently, Bengalese finches produce and perceive their songs in larger perceptual 'chunks' (Suge & Okanoya, 2010). Memory demands can be reduced if, rather than processing transitions between every song note, those notes which always co-occur are instead treated as a single element or chunk. Therefore, memory load can be greatly reduced if the animals learn that the syllables within a 'word' always co-occur and therefore the 'word' could be treated as a single entity (giving four higher-level elements or 'chunks'). This would change the position of the AG structure in the parameter space, both reducing the number of elements and increasing the linearity, making it much simpler to learn (Figure 2.4). Furthermore, it could be argued that the animals could simplify the task further if they could discern that any 'word' can be followed by any other. Here, the structure is completely random and the animals need not monitor the transitions between 'words' as they could simply consider all stimuli as a single stream of words. This would form a single, higher-level entity/chunk consisting of many 'words', each comprised of three syllables. Different responses to any stimulus other than a 'word' could equally support any of these potential strategies. This would change the position an AG takes within the parameter space, and consequently must affect our interpretation of a species' (or individual's, van Heijningen *et al.*, 2009) AGL abilities (Figure 2.4).

A detailed follow up of this triplet segmentation study by the same group (Newport *et al.*, 2004) extended this research to show that these abilities were not limited to recognising statistical relationships between adjacent pairs of CV syllables, but that they also extend to more long distance, non-adjacent syllables. These experiments used similar stimuli to Hauser *et al.* (2001), however the first syllable of a triplet word only predicted the final syllable, and the second syllables varied randomly between words. In 6 different experiments, Newport *et al.*, showed that tamarins appeared to be sensitive to at least some of these long distance statistical relationships, and were able to detect non-words based only on these nonadjacent associations (2004). While presenting a fascinating insight into abilities of nonhuman primates, this result presents further challenges in mapping a parameter space for AGL.

The parameter spaces defined in Figure 2.3 and Figure 2.4 are based on the assumption that animals would learn AG structures based on adjacent relationships between the elements. These relationships are 1st order Markov processes in which only one preceding element is used to predict the next element (Hurford, 2012). However, if the animals can process non-adjacent relationships (higher order Markov processes in which several preceding elements may be used to inform predictions about the next element), such 'rule learning' could simplify having to learn a larger number of pairwise transitions. For example, the A^nB^n structure, which generates the sequence AABB where $n = 2$, appears to be at least partially non-linear when only adjacent elements are considered (e.g. A can be followed by either A or B). However, if an animal is capable of recognising non-adjacent relationships then predictions about the next possible element in a sequence can be made based not only on one preceding element, but on several. It would therefore be possible to form chunk 'AB' if 'A' is always followed by 'B'. In this case, the A^nB^n structure becomes entirely linear (Start → A; if A → AA; if AA → B; if AB → B, if BB → End) and would therefore occupy a different position within the parameter space (Figure 2.4). These examples illustrate that robust empirical data is often required to describe not only which AG structures different animals are able to learn within any given parameter space, but also the strategies they might use to process these structures, which might change where the particular AGL capabilities of a given species are mapped. It may be prudent to take a conservative approach and assume that both human and nonhuman participants would adopt the least cognitively demanding strategy wherever possible. However, empirical testing, possibly with more detailed methods, is required to identify how different species process these AG structures.

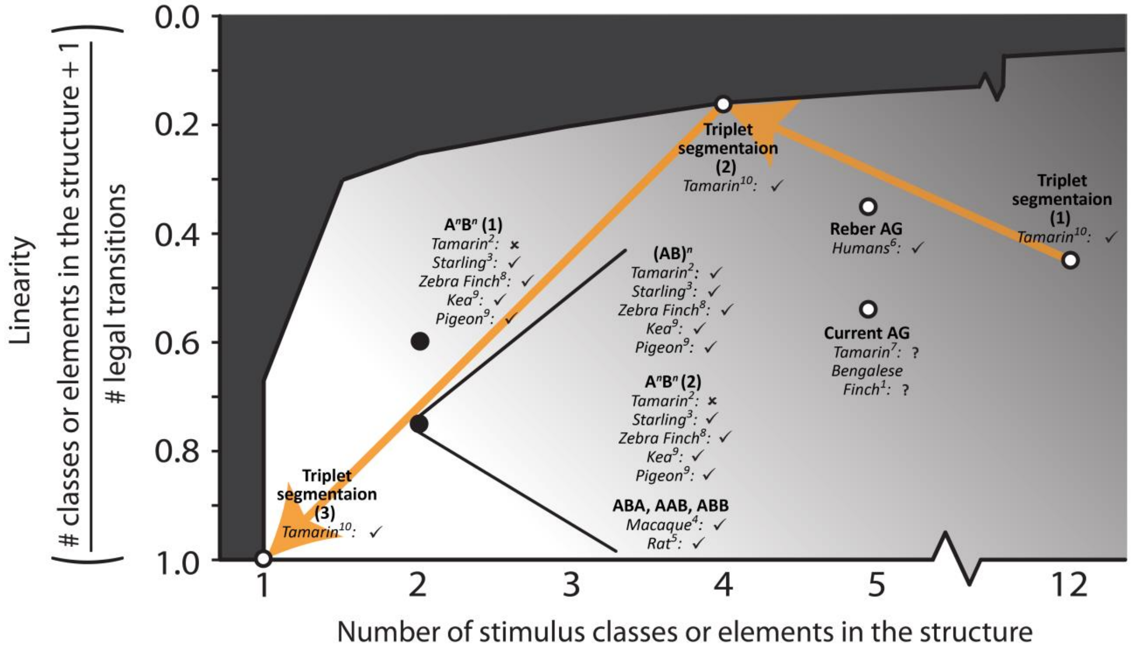


Figure 2.4 Mapping additional AG structures. Shown is a more extensive version of Figure 2.3, where additional AG structures and species' abilities are mapped. Also illustrated is that the mapping of an AG structure is often not fixed because the structure of an AG could be learned in different ways, as follows. Triplet segmentation studies: In 'triplet segmentation' studies such as Hauser *et al.* (2001), participants may represent the 12 unique syllables as independent elements, and learn the relationships between each pair (Triplet segmentation 1). Alternatively, memory load may be reduced by combining the syllables in each word into 4, higher level 'chunks' (Triplet segmentation 2). Furthermore, the words in this experiment could be presented in any order; therefore they could be represented as a still higher category of 'words' with no structure, simply a repetition of different words (Triplet segmentation 3). Adjacent vs. non-adjacent relationships: If the animals can process non-adjacent relationships between distant elements, such 'rule learning' would simplify having to learn a number of pairwise transitions. Based on first order, adjacent relationships, the A^nB^n structure, which generates AABB sequences, appears to be non-linear (A can be followed by either A or B, A^nB^n (1) in the figure). However, if non-adjacent relationships can be learned then the previous one to two elements can be used to predict the following element and the structure would therefore occupy a different position within the parameter space (see A^nB^n (2) in figure). Figure references: 1: Abe & Watanabe, 2011; 2: Fitch & Hauser, 2004; 3: Gentner *et al.*, 2006; 4: Hauser & Glynn, 2009; 5: Murphy *et al.*, 2008; 6: Reber, 1967; 7: Saffran *et al.*, 2008; 8: van Heijningen *et al.*, 2009; 9: Stobbe *et al.*, 2012; 10: Hauser *et al.*, 2001.

2.3. Summary of nonhuman animal AGL capabilities

The parameter space described in Figure 2.3 and Figure 2.4 suggests an objective method to quantitatively compare and evaluate the results from previous AGL studies. While a number of nonhuman AGL experiments have been performed in recent years, large areas of the parameter space remain unexplored, and the strategies supporting different forms of AGL are currently inadequately addressed. Large amounts of evidence have suggested that all species tested are capable of learning relatively linear AG grammars consisting of two stimulus classes (Fitch & Hauser, 2004; Gentner *et al.*, 2006; Murphy *et al.*, 2008; Hauser & Glynn, 2009; van Heijningen *et al.*, 2009; Stobbe *et al.*, 2012). This suggests that all animals are at least sensitive to violations of such invariant AGs structures based on the relationships between adjacent elements, and that they are capable of categorising stimuli based on acoustic properties. There is also convincing evidence that a number of birds are capable of learning less linear AGs of the form A^nB^n (Gentner *et al.*, 2006; van Heijningen *et al.*, 2009; Stobbe *et al.*, 2012). Cotton-top tamarins however proved unable to recognise violations of this grammar, possibly due to an increase in the complexity or non-linearity of the artificial grammar (Fitch & Hauser, 2004). Furthermore, there is evidence that nonhuman primates are also capable of using statistical regularities (Hauser *et al.*, 2001; Newport *et al.*, 2004), to identify word boundaries, although evidence describing how they may process such sequences is currently lacking. Finally, relatively little research has focussed on the more non-linear, non-adjacent AGs with larger numbers of elements which occupy the right half of Figure 2.3. Both songbirds and monkeys have been tested with this type of grammar (Saffran *et al.*, 2008; Abe & Watanabe, 2011), but in both cases experimental issues have made the interpretation of the results difficult (Berwick *et al.*, 2011; Beckers *et al.*, 2012).

Relative to many commonly used AGs, these non-deterministic, forward-branching AG structures depart from the requirement that stimuli are categorised into only two stimulus classes. Rather, several elements, both obligatory and optional, contribute to the structure. Furthermore, while the two stimulus class AGs, after categorisation has occurred, only produce fixed sequences, (e.g. AABB), branching grammars produce a much wider range of sequences from which statistical regularities or rules must be extracted. A number of AGs used to test humans have a forward-branching structure similar to these (e.g. Reber, 1967; Friederici *et al.*, 2002; Udden *et al.*, 2008), and branching structures with varying levels of predictability or linearity can also be observed in the natural song production of several species (Honda & Okanoya, 1999; Okanoya, 2004b; Bolhuis *et al.*, 2010; Berwick *et al.*, 2011;

Hurford, 2012). Furthermore, word transitions in sentences of natural languages are characterised by non-determinism: sentences are not fixed, predetermined sequences, but vary considerably in composition, word transitions and length. Well-formed sentences contain obligatory components (e.g. a subject and a finite verb in English declaratives), as well as varying numbers of optional categories (adjectives, adverbs, etc.), the positions of which depend on the other words in the sentence. Language learners must deal with unpredictable variation (Hudson Kam & Newport, 2009) and appear to have a general bias to reduce such variation during learning (Smith & Wonnacott, 2010). Thus, the ability to process and learn from such unpredictable variation may also be central to human language processing. Therefore rigorous, comparative assessment of these abilities in nonhuman primates represents an important opportunity to investigate the extent to which these abilities might be unique to humans or may be more generic and common to other nonhuman primates. In the next section, I will propose hypotheses regarding how nonhuman primates may learn these forward-branching, non-deterministic AG structures, and how this might inform us about aspects of language processing in humans.

2.4. Plans and hypotheses

Forward-branching, non-deterministic artificial grammars present the opportunity to explore how nonhuman primates might learn AG structures that can generate a wide variety of different sequences, and how they use this information to identify sequences that violate these structures. These abilities, to extract rules or statistical relationships from a varied and complex input, are crucial to the learning and processing of natural language (Saffran, 2002; Saffran *et al.*, 2008). However, to provide meaningful insights into the abilities of nonhuman animals, the evolution of these abilities, or to develop animal models to investigate the neurological mechanisms that support them, a number of empirical questions must be answered. Are nonhuman primates able to learn AGs of this form and complexity, and if so, might this ability be based purely on the use of simple cues (such as the familiarity of the sequences in the experiment performed by Saffran *et al.*, 2008)? How might these abilities vary between different species, and how can this inform us about the evolution of such capabilities? If nonhuman animals are able to learn these AG structures, do the cognitive mechanisms and brain areas recruited reflect those used in humans in the processing of natural languages or artificial grammar learning? It is possible that comparable behavioural results in different species might be supported be distinct strategies and different networks of brain areas, therefore it is critical to explore whether any abilities seen in monkeys might represent comparable processes in humans.

In this thesis I will present a series of experiments that aim to address these questions. In Chapter 3, I will describe two experiments in Rhesus macaques investigating, both at the group level and more objectively in individual animals, whether these Old World monkeys are able to learn the structure of these AGs and how they may recognise sequences that violate the AG structure. In Chapter 3, I will extend these experiments to assess the abilities of human participants and common marmosets, a species of New World monkey more closely related to the tamarins tested in previous AGL experiments. These experiments may be used to test the hypothesis that the abilities observed in macaques might be, to some extent, conserved in other more distantly related primate species. Finally, in Chapter 5, I will discuss fMRI experiments in both Rhesus macaques and human participants, to assess the brain areas involved in AGL. This evidence could either support or refute the hypothesis that comparable abilities observed between species would be supported by a broadly similar network of brain areas and that these areas that may be involved in language in humans. The goal of this thesis is to provide empirical evidence assessing whether nonhuman primates may be able to perform AGL tasks of sufficient complexity to inform us about aspects of language in humans, and how they might process them. This research has the potential to both inform us about how the capacity for language may have evolved in humans and to identify potential candidate animal model systems in which the mechanisms supporting these abilities can be explored at a neuronal level which is not possible in human participants.

Chapter 3: Behavioural Evidence for AGL in Rhesus Macaques

Parts of this chapter have been published as:

B. Wilson, H. Slater, Y. Kikuchi, A. E. Milne, W. D. Marslen-Wilson, K. Smith & C. I. Petkov (2012). Auditory artificial grammar learning in macaque and marmoset monkeys. *The Journal of Neuroscience,* **33(48)**, 18825-18835.

All work presented in this chapter was primarily conducted by the author, except where explicitly stated. Co-authors on this paper provided valuable comments and discussion, and H. Slater and A. Milne provided additional ratings of the macaques' responses in the video-coding experiment. The majority of the experimental design, the running of the experiments, data analysis and writing were conducted by the first author.

3.1. Abstract

Artificial Grammars (AG) can be designed to emulate aspects of human language, and AG Learning (AGL) paradigms used to study the extent of nonhuman animals' structure-learning capabilities. These approaches could provide insights on the evolution of human syntactic abilities. Previous quantitative comparisons of different AG structures (see Chapter 2) highlighted a gap in our understanding of nonhuman primate AGL abilities. To address this, I tested whether Rhesus macaques (*Macaca mulatta*) could learn an AG with a prominent forward-branching structure, designed to emulate certain aspects of the non-deterministic nature of word transitions in sentences or elaborations in animal songs. Initially, 13 macaques were habituated to sequences of nonsense words generated by the AG structure. The animals were then presented with sequences that either followed the AG ('consistent') or that violated the structure ('violation') from a concealed audio speaker. The macaques showed greater dishabituation responses to the violation sequences. Furthermore, these responses could not be attributed to either very simple learning strategies or rote memorisation of the habituation sequences. A second experiment used a novel eye-tracking approach to objectively explore AGL in three individual macaques. This method provided higher levels of temporal precision and confirmed the results of the first experiment in individual animals, and further ruled out simple learning strategies. These experiments provide evidence for a previously unknown level of complexity in the AGL capabilities of an Old World monkey, suggesting that some sequencing abilities critical for human syntax may be evolutionarily conserved and thus not uniquely human or language specific.

3.2. Introduction

Language is a uniquely human trait with poorly understood evolutionary origins (Bickerton & Szathmary, 2009; Hurford, 2012). Owing to its complexity in meaning ('semantics') and structure ('syntax'), natural language cannot be directly studied in nonhuman animals. However, theoretical work has identified distinct computations related to language that could be comparatively studied (Fitch, 2000; Hauser *et al.*, 2002; Bickerton & Szathmary, 2009; Hurford, 2012). Initial approaches studied referential communication in animals (Seyfarth *et al.*, 1980; Fitch, 2000), which has inspired work on how neurons process communication signals (Ghazanfar & Hauser, 2001; Tian *et al.*, 2001; Gifford *et al.*, 2005). Recently, songbirds have been viewed as promising neurobiological model systems because, like humans and a few other animal species, they are vocal learners and can produce songs with certain 'syntax-like' qualities (Jarvis, 2004; Berwick *et al.*, 2011; Petkov & Wilson, 2012). Yet, vocal learning appears to have occurred by convergent evolution rather than by common descent, since nonhuman primates and most other species have more limited vocal production capabilities (Jarvis, 2004; Petkov & Wilson, 2012). This has raised questions regarding whether nonhuman primates might be able to learn structures or patterns with sufficient levels of complexity to provide novel insights into possible language precursors in extant species. The discovery of such abilities has the potential to inform us about the evolution of language related abilities in humans, and represents the necessary first step in developing a nonhuman primate model system to study syntactic precursors at the neuronal level (Hauser *et al.*, 2002; Fitch & Friederici, 2012).

Artificial Grammars (AG) can be created to emulate certain aspects of the structure of natural language or simpler "rule-based" structures that some animals might be able to learn. These can be comparatively studied using AG Learning (AGL) paradigms (e.g. Fitch & Hauser, 2004). For example, many studies have used AGL to investigate how nonhuman animals, human infants or adults acquire the statistical relationships between sensory stimuli in sequences generated by an AG, referred to as 'statistical learning' (Saffran *et al.*, 1996b; Saffran *et al.*, 1999). In such studies the participant has no *a priori* knowledge about the structure of the AG. Yet, by being habituated to or trained with exemplary sequences generated from the AG, the relationship between the elements in the sequence could be acquired. Differential responses to novel well-formed (consistent) sequences compared to those that violate the AG structure suggest that some aspect of the AG structure was learned. Although several nonhuman animal AGL studies have been conducted (e.g. Fitch & Hauser, 2004; Gentner *et al.*, 2006; Saffran *et al.*, 2008), the AGL capabilities of Old

World monkeys, which are more closely related to humans than any other species in which AGL experiments are conducted, have rarely been tested. The only study in Old World monkeys habituated free ranging Rhesus macaques to sequences of macaque vocalisations of the form AAB (e.g. 'coo', 'coo', 'grunt'), and found stronger dishabituation responses to ABB sequences (e.g., 'coo', 'grunt', 'grunt', Hauser & Glynn, 2009). While it is possible that the animals' learning was based on whether the first element of the sequence was repeated, these results at least suggest that Rhesus macaques naturally show some sensitivity to the temporal order of elements in a sequence and thus demonstrate some AGL abilities. Furthermore, even beyond the Old World monkeys, there is currently limited evidence that any nonhuman animal can learn the structure of non-deterministic, branching AGs that produce more varied sequences (see Chapter 2, Saffran *et al.*, 2008; Abe & Watanabe, 2011; Berwick *et al.*, 2011; Beckers *et al.*, 2012).

In two experiments, I aimed to investigate the abilities of Rhesus macaques (*Macaca mulatta*) to learn a forward-branching AG. Learning this non-deterministic structure, requires the animals to recognise statistical relationships between a number of different nonsense word elements which can occur in a wide variety of positions and arrangements within sequences of varying composition and length. While there is strong evidence that human adults and infants (Saffran, 2002; Saffran *et al.*, 2008) can learn such grammars, as discussed previously (see Chapter 2), the abilities of nonhuman animals appear to be less clear (Berwick *et al.*, 2011; Beckers *et al.*, 2012; Hurford, 2012). After cotton-top tamarins (*Saguinus oedipus*) were habituated to the AG structure, the only evidence for significant dishabituation responses to violations of the AG structure was obtained when the animals were tested with the same 'consistent' sequences to which they had been habituated (Saffran *et al.*, 2008). Thus the animals' dishabituation responses may have been based primarily on the monkeys being unfamiliar with the violation test sequences. The experimental design of the current experiments aimed to assess the abilities of Rhesus macaques while avoiding these potential experimental confounds by testing the monkeys with both 'familiar' and 'novel' consistent (well-formed) testing sequences to determine whether macaques would primarily distinguish between sequences on the basis of familiarity (Figure 3.1 and Methods). Secondly, in a study testing Bengalese finches on a similar AG structure (Abe & Watanabe, 2011), several authors have noted that the testing sequences differed significantly in their acoustic properties between conditions (Beckers *et al.*, 2012; Berwick *et al.*, 2012). All consistent sequences were acoustically very similar to each other but the violation sequences differed considerably, thus, the animals could have responded differentially to the test sequences based solely on acoustic differences. Therefore, to explore the extent to which the animals learned the sequences, the testing sequences in the current experiment included violation sequences that violate the AG structure at multiple positions, and allowed control for acoustic differences between consistent and violation sequences (Figure 3.1). Finally, it has been noted that many nonhuman animal AGL studies fail to clarify what parts of the sequence animals monitor for violations (van Heijningen *et al.*, 2009; ten Cate & Okanoya, 2012). For example, in experiments which present animals with test sequences of the form AABB and ABAB (Fitch & Hauser, 2004; Gentner *et al.*, 2006), the animals need only detect violations in the first two elements of the sequences. Thus, this study incorporated two different types of violation sequences containing very early violations (at the first element) and those with only later violations (Figure 3.1). Furthermore, additional analyses, using a novel eye-tracking paradigm provided the temporal precision needed to determine how the animals evaluated specific elements throughout the test sequences. Different patterns of responses to these stimuli have the potential to inform us about the various learning strategies the animals may use. For example, stronger responses to 'novel' compared to 'familiar' sequences may suggest that, like the tamarins, macaques are sensitive to sequences of sounds with which they are not familiar. Similarly, larger responses to sequences that 'do not begin with A' compared to the violation sequences that 'begin with A' show that animals recognise violations in the first element of the sequence and use this to cue their responses. Finally, the key contrast is between 'novel' consistent sequences and violation sequences that 'begin with A', because differences in responses to these sequences cannot be attributed to either the novelty or familiarity of the sequence, or containing an obvious, initial violation, therefore reflect strong evidence of artificial grammar learning.

Two experiments were designed to assess macaques' abilities to learn this AG. The first used a modified version of the traditional video-coding method. The natural responses of animals to test sequences presented from a concealed speaker were video-recorded for later offline analysis by multiple raters. This approach provides population level data to investigate AGL abilities within a group of animals. Although every effort was made to standardise the video-coding procedure and ensure high levels of objectivity and consistency between raters, this approach carries an inherent level of subjectivity. Therefore, I developed a novel eye-tracking paradigm to objectively measure the looking responses of individual macaques. This approach provided far greater temporal precision and statistical power than the traditional video-coding approach, allowing fine-grained analyses of responses to specific test sequences in individual animals, as is required for neuroscientific experiments (e.g. fMRI, EEG, electrophysiological recordings). The video-coding experiment aimed to provide population level data on macaque AGL and to validate the use of the eye-tracking method, which in turn is capable of providing more detailed behavioural data than any previous nonhuman primate AGL task.

3.3. General methods

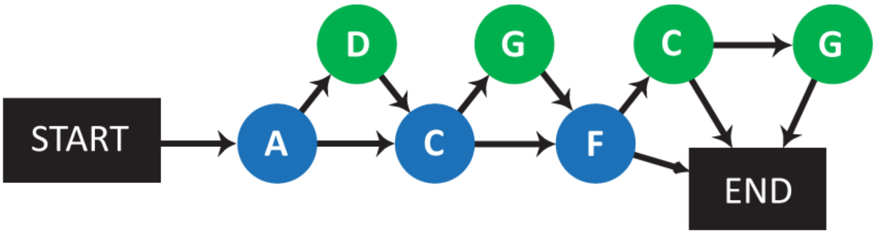
3.3.1. Stimuli

The video-coding and eye-tracking experiments both used identical stimuli. The artificial grammar used in this experiment was based on that of Saffran (2002; 2008) and is described in Figure 3.1. Five different elements (nonsense words) can be combined in a variety of orders to form 'consistent' legal sequences by following the path of arrows through the AG. In this way both 'familiar' test sequences (which also appear in the habituation phase) and 'novel' sequences which had not previously been presented were generated. 'Violation' sequences were created by designing sequences that could not be made by following the arrows from Start to End. Half of the violation sequences, like the consistent sequences, began with the A element, while the other half did not begin with A, thus violation the AG structure from the earliest possible element.

The nonsense words were produced by a female native-English speaker and recorded with an Edirol R-09HR (Roland Corp.) sound recorder. The amplitude of the recorded sounds was root-mean-square (RMS) balanced and the nonsense word stimuli were combined into sequences (100ms inter-stimulus intervals, ISI) using customised Matlab scripts. The sounds were presented to the animals using Cortex software (Salk Institute) at ~75 dB SPL (calibrated with an XL2 sound level meter, NTI Audio).

The duration of the naturally spoken nonsense word stimuli within the sequences varied (Klor = 0.64 sec; Jux = 0.62 sec; Cav = 0.56 sec; Biff = 0.40 sec; Dupp = 0.39 sec). Learning of any artificial grammar requires the ability to discriminate between the elements in a sequence to learn the relationships between them, and duration differences between the nonsense word stimuli provide one additional cue. However, to demonstrate that the AG has occurred, it is necessary to show that simple duration cues could not be used to identify consistent or violation sequences. I confirmed that the duration of the sequences could not be used as a cue to sequence condition, as follows. The consistent and violation sequence sets were balanced in the number of elements in the sequences (Figure 3.1) and the mean length (and SD) of the sequences were comparable: consistent sequences, 3.14 (0.42) sec; violation sequences, 3.25 (0.28) sec. Furthermore, I confirmed that there was no significant difference in sequence duration between consistent and violation sequences (independent samples t-test, $t_6 = 0.435$, $p = 0.68$), or in the duration of the individual elements present in consistent vs. violation sequences ($t_{42} = 0.609$, $p = 0.55$).

Further steps were taken in designing the sequence sets to balance for acoustical differences, by either balancing for the presence of the different elements (A, C, D, F, G), as much as possible, or analytically confirming that acoustical differences could not explain the reported results. The A, F and G elements were balanced so that they occurred equally often in each of the consistent and violation sequences (Figure 3.1). Half of the violation and consistent sequences were also balanced for the presence of the C and D elements, but it was difficult to achieve this balance in the other half of the sequences without introducing other potential confounds. Nonetheless, acoustical differences cannot explain the results for the following reasons. First, eye-tracking results by acoustical element (Figure 3.7) showed that a comparable pattern of stronger responses to elements in violation vs. consistent sequences were made in response to all of the elements. Therefore the monkeys did not simply respond strongly to certain elements, but their responses varied based on the context in which the element occurred (consistent or violation). Second, an analysis of the average eye position in response to the C and D elements (ANOVA factors: element (C or D), condition (consistent or violation) and monkey) showed the expected main effect of condition ($p = 0.001$) and monkey ($p = 0.008$), but no effect was observed for the element factor ($p = 0.13$) and no interactions were seen between the elements and condition or monkey (all $p > 0.1$). Therefore, the responses cannot be explained by a preference for any acoustical element, but only by the context in which the element occurs.



Nonsense words

A	‘klor’
C	‘biff’
D	‘jux’
F	‘cav’
G	‘dupp’

Habituation sequences
ACF
ACFC
ACGF
ACGFC
ACGFCG
ADCF
ADCFC
ADCFCG
ADCGF

Testing sequences		
ACGFC	Consistent	Familiar
ADCFCG	Consistent	Familiar
ACFCG	Consistent	Novel
ADCGFC	Consistent	Novel
AFGCD	Violation	Begin with A
AFCDGC	Violation	Begin with A
FADGC	Violation	Do not begin with A
DCAFGC	Violation	Do not begin with A

Figure 3.1. Artificial Grammar structure. The AG structure contains five unique elements and multiple forward-branching relationships. Consistent sequences (strings of nonsense words) are generated by following any path of arrows from START to END. Violation sequences do not follow the arrows. The AG was used to create nine habituation sequences, which included all possible legal transitions between elements. All experiments began with habituating the animals to the habituation sequences, presented in a randomised order. In the following testing phase, animals were presented with the 8 testing sequences (including ‘familiar’ and ‘novel’ consistent sequences, and violation sequences which ‘begin with A’ or ‘do not begin with A’) in a randomised order.

The statistical relationships between pairs of elements in an AG can be quantified by calculating the transitional probability of the relationship, i.e. the probability that two specific elements will co-occur. Transitional probability of element x immediately preceding element y is given by:

$$TP_{xy} = p(y|x) = \frac{p(xy)}{p(x)}$$

Where a TP of 1.0 means x will always be followed by y and a TP of 0 means x will never be followed by y. Using this approach, the transitional probabilities between all pairs of elements in the habituation sequences were calculated (Table 3.1). This highlights the variability with the AG, with some elements able to occur in more legal positions than others and different transitions being more or less predictable. For example, 'C' can be followed by either 'F', 'G' or 'End', while 'D' can only ever be followed by 'C'. To learn the AG the animals must extract this statistical information from the habituation sequences, and cannot simply rely on learning a fixed pattern, such as the stimuli will always alternate, as in the $(AB)^n$ structure. The average TP for each test sequence was also calculated (the mean of the TPs between each pair of elements in the sequence). This demonstrates that the consistent sequences, which contain only legal transitions, have relatively high TPs. The violation sequences however, which contain illegal transitions that, by definition, have a TP of 0, have much lower average TPs. Therefore, if the animals are sensitive to these statistical relationships, it is possible to discriminate between the consistent and violation sequences.

Table 3.1 Transitional probabilities between pairs of elements and in test sequences. The transitional probability (TP) of every legal transition between elements was calculated based on the frequency of their occurrence within the habituation sequences. Higher TPs represent more common transitions. The average TP of each test sequence is also shown, highlighting the higher average TPs in the consistent than the violation sequences.

Transition	Transitional Probability (TP)
Start – A	1.00
A – C	0.56
A – D	0.44
D – C	1.00
C – F	0.36
C – G	0.43
C – End	0.21
F – C	0.56
F – End	0.44
G – F	0.67
G – End	0.33

Test Sequences	Average TP
Consistent	
ACGFC	0.57
ADCFCG	0.59
ACFCG	0.54
ADCGFC	0.62
Violation	
AFGCD	0.17
AFCDGC	0.25
FADGC	0.11
DCAFGC	0.17

3.4. Video-coding experiment

3.4.1. Methods

Participants

Thirteen male Rhesus macaques (*Macaca mulatta*) participated in this experiment. The macaques were housed in two separate group-housed colonies and were tested on two separate days. The animals were individually housed for testing, wherever possible.

Ethics statement

All animal work and procedures were approved by the U.K. Home Office and abide by the Animal Scientific Procedures Act (1986) on the care and use of animals in research, which promotes the principles of the 3Rs: replacement, reduction and refinement of research work with animals. This research study in nonhuman primates abides by the recommendations of the Weatherall report on "The use of nonhuman primates in research". Moreover, the laboratory work has been peer reviewed by anonymous referees, and the National Centre for 3Rs before being funded by the Wellcome Trust to conduct this research. Lastly, this project is a part of an effort to underpin behavioural capabilities in nonhuman primates that could be studied at the neuronal level, which is difficult to do in humans. The most appropriate species for this project have also been considered and selected prior to the start of the project.

Habituation phase

During the habituation phase, the animals were presented with the habituation sequences in a randomised order (Figure 3.1). The sequences were presented from an audio speaker concealed within the colony (rate of 9 sequences/min; inter-sequence interval = 4 sec). Habituation occurred for two hours on the afternoon prior to the experiment, when the animals were quiet and relaxed, but a few hours before the lights would be turned off for them to sleep. The following morning the animals were re-habituated to the sequences presented in a randomised order for 10 minutes, immediately prior to the start of the experiment.

Testing phase

Video cameras were set up early in the morning to allow the animals to become habituated to their presence. During testing, a randomly selected test sequence of the 8 (consistent or violation) sequences (Fig. 1B) was individually presented (4 times each, for a total of 32 testing trials; at an average rate of 1/min; inter-sequence intervals ranged between 45 and 75 sec). Each animal's orienting responses were video recorded for offline analysis (JVC and Sony digital video cameras; 720 x 576 resolution; 25 frames/s).

Video-coding Procedure

The traditional video-coding procedure (e.g. Hauser *et al.*, 2001; Fitch & Hauser, 2004; Newport *et al.*, 2004; Saffran *et al.*, 2008) was refined to minimise subjectivity in video-coding analysis. First, the audio track for each video was digitally scrambled so that it was not possible to identify the sequence condition. Three independent raters blind coded all of the videos from each animal. Each rater coded orienting responses and response durations based on eye, head and/or body movements in the direction of the concealed audio speaker that presented the stimulus sequences. The strength of the orienting responses were recorded on a five point Likert scale, 1 = no orienting response; 2 = probably no response; 3 = ambiguous response; 4 = probable orienting response; 5 = definite orienting response.

Data analysis

For all analyses only the trials on which the majority of the raters (2 out of 3) agreed that an unambiguous response was made were used (strength of response ≥ 4, on the scale of 1-5). The proportion of trials on which the animals unambiguously responded, the strength and the duration of these responses were analysed. Analyses based on the duration or strength of responses included only those trials on which a response was recorded to ensure that these results were not biased by a higher number of responses in one experimental condition.

Inter-rater reliability

Three raters coded all of the videos. Inter-rater reliability was calculated pairwise between the raters. In the macaque experiment, the raters, on average, had exact agreement on the strength of the response (on the five point scale) on 75.4% of the trials and were within one response point from each other on 85.1% of the trials. Cohen's Kappa revealed "substantial" average agreement between raters, $K = 0.67$ (Landis & Koch, 1977). The macaques were rated as unambiguously responding to 14.7% of all recorded trials by a majority of raters resulting in 16 grammatical and 45 ungrammatical response trials (total of 61) used for analysis.

3.4.2. Results

The 13 macaques showed a significantly higher proportion of orienting responses to the violation than the consistent sequences (paired samples *t*-test, $t_{12} = 7.898$, $p < 0.001$;

Figure 3.2A). The two different types of consistent and violation sequences were analysed to clarify whether the observed effect depends on either familiarisation or the animals only noticing violations in the first sequence element (i.e., sequences which, unlike the consistent sequences, 'do not begin with A'; Figure 3.2B). A repeated measures analysis of variance (RM-ANOVA) with four levels of the factor: sequence condition ('familiar', 'novel', 'begins with A' and 'does not begin with A') revealed a strong main effect of sequence condition ($F_{3,36} = 9.146$, $p < 0.001$; Figure 3.2B). Bonferroni post-hoc tests showed significant differences between several key contrasts. Most notably, the macaques responded significantly more frequently to the violation sequences that 'begin with A' than the 'novel' consistent ($p < 0.05$; Figure 3.2B), suggesting the macaques' AGL cannot be interpreted based on familiarity or only attending to the first element in the sequence. Further significant differences were observed between 'familiar' sequences and both violation sequences which 'begin with A' and 'do not begin with A' (Bonferroni corrected, $p < 0.05$ Figure 3.2B). No differences were observed between 'familiar' and 'novel' consistent sequences or violation sequences which 'begin with A' and 'do not begin with A', even when more liberal LSD corrections were applied ($p > 0.49$ in both cases). These results suggest that not only did the macaques respond to violations of the AG more frequently, but also that their responses cannot be attributed only to superficial differences between the sequences, such as familiarity or monitoring only the initial parts of the sequences.

A similar, albeit weaker, pattern of effects was observed in analyse of the duration of responses (consistent and violation sequences: $t_{12} = 2.330$, $p = 0.038$; RM-ANOVA 'condition' factor with 4 levels: $F_{3,36} = 5.276$, $p = 0.004$; Figure 3.2C-D). No post-hoc tests between sub-categories survived Bonferroni corrections ($p > 0.05$). As in the analyses of proportion of responsive trials, LSD corrections reveal significant differences between conditions (i.e. 'familiar' and 'begins with A'; 'familiar' and 'does not begin with A'; 'novel' and 'begins with A'; 'novel' and 'does not begin with A'; $p < 0.05$) but not within conditions (i.e. 'familiar' and 'novel'; 'begins with A' and 'does not begin with A'; $p > 0.46$). Therefore, like the proportion of responses, the duration for which the animals respond cannot be attributed to familiarity or attending only to the initial element in the sequence.

Finally, analysis of the strength of responses (based on the video-coders' ratings from 1-5, see Methods), showed that the macaques also responded more strongly to the violation sequences ($t_{12} = 3.129$, $p = 0.009$; RM-ANOVA with the factor: 'condition' with 4 levels: $F_{3,36} = 5.685$, $p = 0.003$; Figure 3.2E-F). Bonferroni post-hoc tests revealed a significant difference between ‘novel’ consistent sequences and violation sequences which ‘begin with A’, but no other contrasts. Again, even LSD corrected post-hoc tests revealed significant differences between all conditions ($p < 0.05$) except ‘familiar’ and ‘novel’ ($p = 0.98$) and ‘begins with A’ and ‘does not begin with A’ ($p = 0.39$).

These results, based on three different measures, suggest that Rhesus macaques are sensitive to violations of an implicitly learned AG structure. No differences in responses were observed between ‘familiar’ and ‘novel’ test sequences in any analysis, even when using liberal methods of multiple comparison correction. Therefore these responses cannot be attributed to familiarity with the test sequences relative to the violation sequences. Furthermore, the animals do not appear to base their responses on violations in the initial sequences positions (based on no differences between sequences which ‘begin with A’ and those which ‘do not begin with A’), despite the potential salience of this cue. Rather, the animals appear to respond to more subtle cues to grammaticality indicating considerable learning of the AG structure. To explore these effects in more detail, and in individual animals, a novel eye-tracking paradigm was developed.

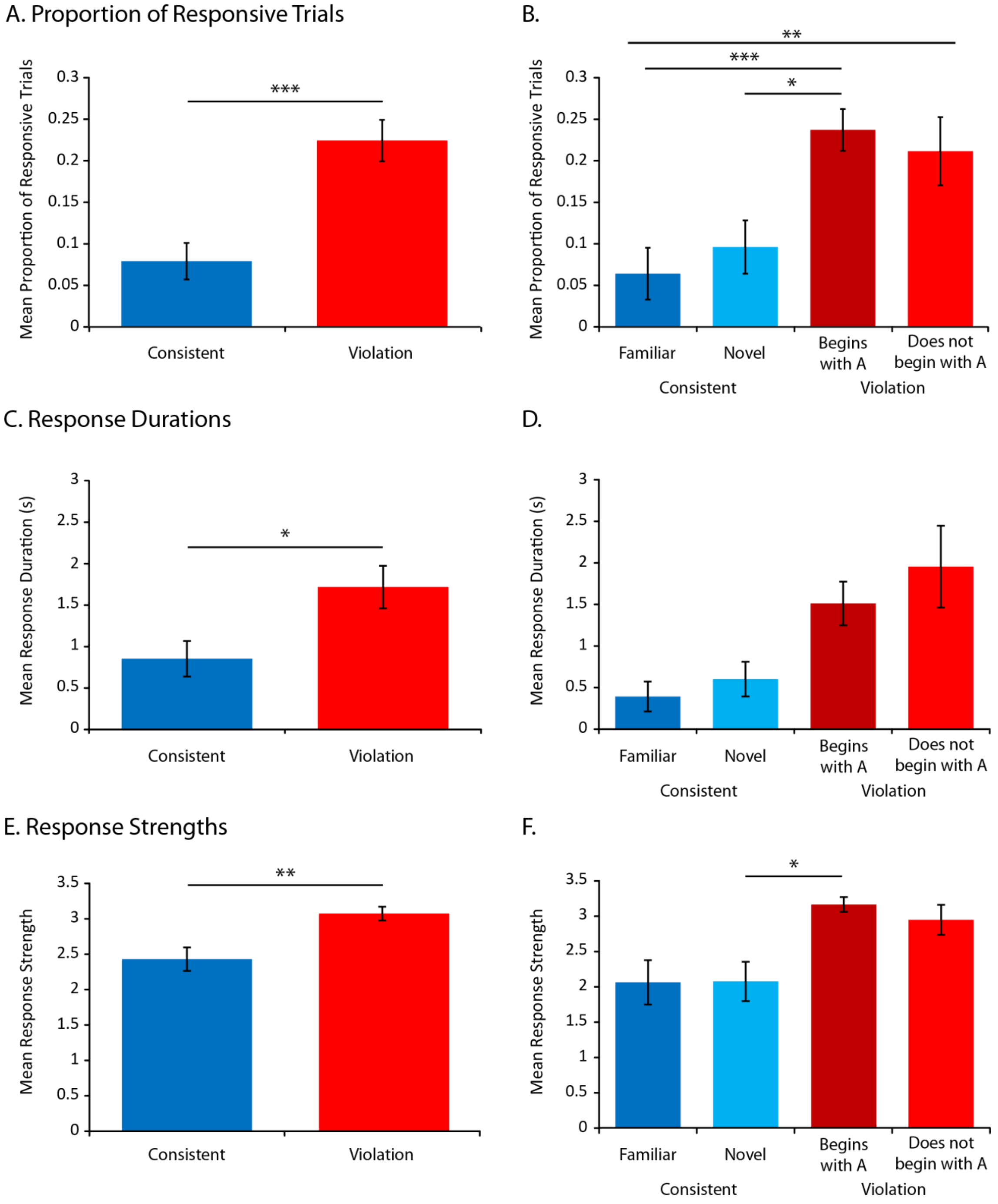


Figure 3.2. Video-coding experiment results in Rhesus macaques. (A-B) Mean proportion (±SEM) of trials on which macaques made unambiguous looking-responses as evaluated by a majority of raters (see Methods). The left panels indicate responses to consistent and violation sequences, right panels display results to specific subsets of stimulus sequences. (C-D) Mean response duration of responses (±SEM) to consistent and violation sequences (C) and subsets of sequences (D). (E-F) Mean responses strengths based on coding from three raters (±SEM) to consistent and violation sequences (E) and subsets of sequences (F). Significance levels are shown for all main effects (paired sample *t*-tests; left panels) and Bonferroni post-hoc tests (right panels); * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

3.5. Eye-tracking experiment

The video-coding experiment used three different response measures to demonstrate, at the group level, that Rhesus macaques appear sensitive to violations of this non-deterministic artificial grammar. However, despite significant effort to improve and control the traditional video-coding procedure, an inherent element of subjectivity remains. Furthermore, although more response measures were reported than are typical in nonhuman AGL studies, they all lack the temporal sensitivity to investigate animals' responses at different points throughout the sequences, for example to specific elements or rule violations. Finally, while demonstrating AGL in Rhesus macaques at the group level is an important and novel result, neuroimaging experiments require confirmation of such abilities in individual animals. Therefore, I developed a novel, objective eye-tracking paradigm to investigate the AGL capabilities of Rhesus macaques with greater sensitivity than has been possible with existing techniques.

3.5.1. Methods

Participants

Three adult male Rhesus macaques (*Maccaca mulatta*) participated in this experiment. All animal work and procedures were approved by the U.K. Home Office and follow the Animal Scientific Procedures Act 1986 on the care and use of animals in research. Animals were selected for this experiment based on training to participate in a visual fixation task to obtain a reward. All three monkeys had participated in the previous video-coding experiment.

Stimuli

The stimuli sequences were identical to those used in the video-coding experiment (Figure 3.1)

Procedure

Animals were seated, with their heads immobilised, in a primate chair 60 cm in front of a computer monitor, and two audio speakers (Creative Inspire T10) horizontally positioned at $\pm 30^{\circ}$ visual angle. Following 25% of successful fixation trials a stimulus sequence was presented from either the left or the right audio speaker, and eye-tracking data was recorded (Figure 3.3A).

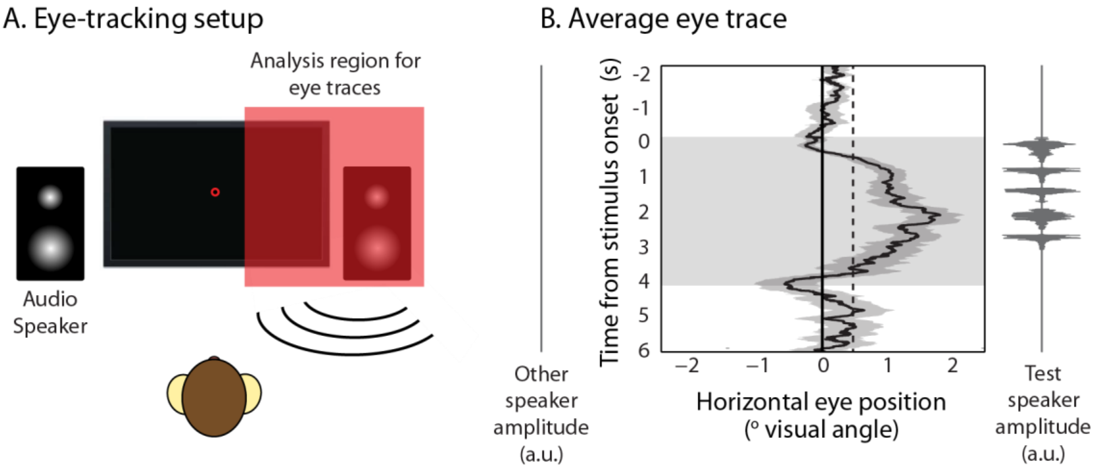


Figure 3.3 Schematic of macaque eye-tracking experiment. (A) Macaques were seated in front of a computer monitor. After fixating on a central fixation spot, on 25% of trials a randomly selected test sequence was presented from one of two audio speakers and eye-tracking responses were recorded. (B) Average eye trace from one monkey (±standard error; SEM). Positive values on the horizontal axis indicate eye movements toward the audio speaker (left or right) that presented a given test sequence. The dotted line denotes 3 SDs of the variance in eye position during fixation, which was used for analysis of significant looking-responses (shaded area is the individually defined response period). Abbreviation: a.u., arbitrary units.

Habituation phase

Each session began with a habituation phase, during which the animal listened to the habituation sequences (Figure 3.1) presented in a random order from both left and right audio speakers for 30 minutes (rate of presentation: 9 sequences/min; inter-sequence interval = 4 sec).

Testing phase

Following the habituation phase was a testing run consisting of multiple trials. Each trial began when the animal engaged a red fixation circle in the centre of the screen to centre the eyes. If the animal continuously fixated for two seconds it was given a juice reward, and 25% of the successful fixation trials were followed by a testing trial in which a randomly selected testing sequence (of the 8 possible, Figure 3.1) was randomly presented from either the right or the left audio speaker. The trials on which a testing sequence was presented were separated by on average 4 trials where no test sequence was presented and the animal only fixated. Experimental data were collected in 1-5 separate testing runs per day. Each testing run included at least eight trials (one presentation of a randomised order of every test sequence, Figure 3.1). The duration of a testing run was dictated by the animal's motivation to continue engaging in the fixation task for a juice reward. However, data analysis was only conducted on the first 8 trials (1 presentation of each of the testing sequences), because all animals completed these, thus, each animal was presented with each of the 8 testing sequences either in each testing run (25 or 26 runs per animal). Furthermore, while monkeys showed consistent levels of looking responses throughout these testing runs (Figure 3.4A), they only appeared to discriminate between the consistent and violation sequences for the first 8 sequences of any testing run (Figure 3.4B). The animal was given a short break between each testing run, during which the animal listened to a new randomised set of habituation sequences for five minutes in order to re-habituate him to the AG structure. After this, another testing run began, if the animal remained motivated to fixate to start each trial.

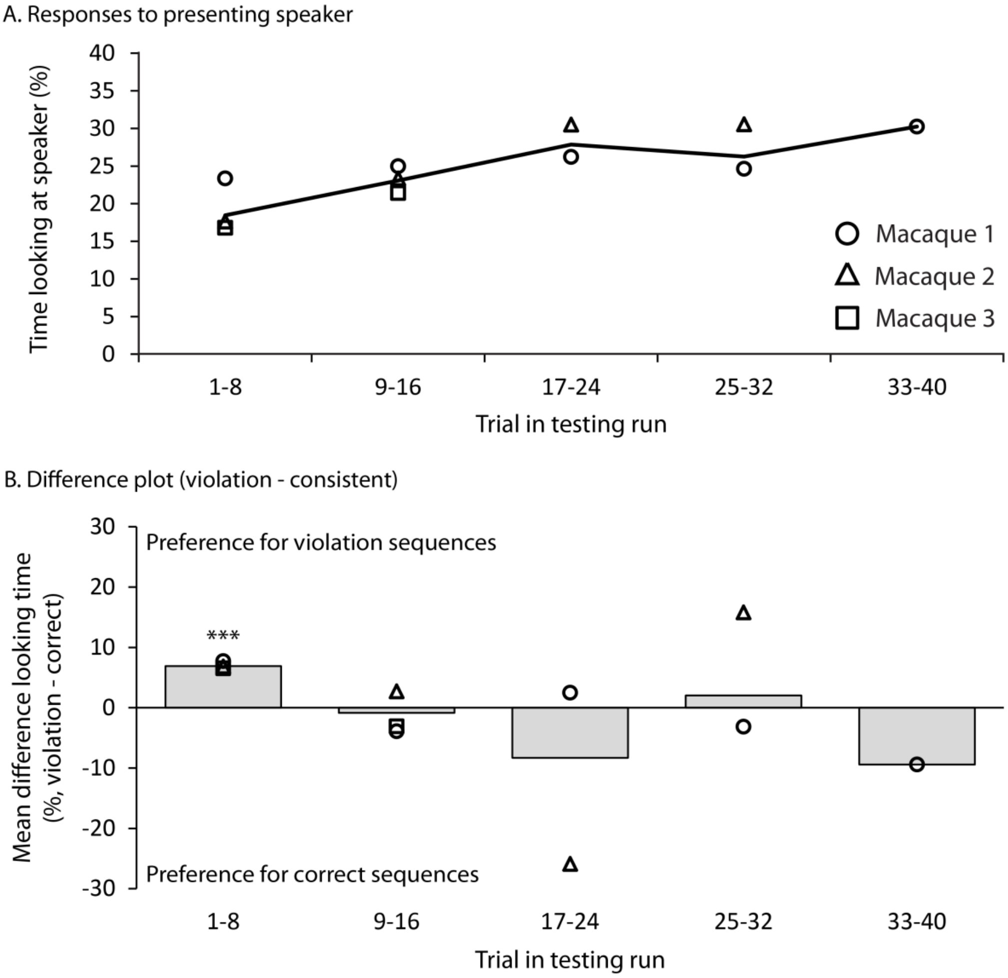


Figure 3.4 Eye-tracking responses across testing runs. (A) Duration of responses throughout the testing run (% time looking at speaker with average individual animal identified by symbols). Each testing run was divided into bins containing 8 stimulus trials. Plotting the average response duration of these trials, regardless of stimulus condition, shows that the animals continued to look in the direction of the presenting speaker throughout the testing run and did not appear to lose interest in looking at the test sequences (when they were motivated to participate and data could be obtained). (B) Shown are the differences in response durations between the violation and consistent sequences for each bin of eight stimulus trials (% time looking at speaker, with symbols identifying average values for individual animals). Looking durations were significantly higher in response to violation sequences compared to consistent sequences in the first eight trials ($t_{63} = 4.429$, $p < 0.001$). After the first bin no difference was observed between the violation and consistent sequences (in all cases, $p > 0.1$). These results suggest that the effects of habituation (which occurred before the start of each testing run in the form of either a habituation or re-familiarisation period) do not persist indefinitely and that to maintain the effect shorter runs and more frequent periods of re-familiarisation are needed. It is important to note that the increase in individual variability observed in later testing runs is due to reduced statistical power, since (although all animals completed at least the first eight testing trials) not all animals managed to complete all of the subsequent testing trials in each run.

Data analysis

The three macaques participated in 25, 25 and 26 testing runs respectively. Only the first eight trials of each testing run were used for further analysis since all of the animals completed these (see Figure 3.4). The eye-tracking data for each trial contained both the 2 second baseline period during which the animal fixated on the central fixation spot and the subsequent period during which the test sequence was presented randomly from one of the two audio speakers Figure 3.3B. Looking-responses to the test sequences were defined individually for each animal as looks towards the presenting audio speaker (left or right) exceeding 3 SDs of the variability in the baseline eye fixation period. The analysis included the time from stimulus onset up to the point when the animal looked in the opposite direction for more than 200ms. This identified when the animal seemed to lose interest in the test sequence and looked above 3 SDs of baseline variability towards the opposite, silent audio speaker (Figure 3.3B). The length of the response window for the three monkeys (M) was: M1 = 2128ms, M2 = 2984ms, M3 = 4180ms. The data were also analysed using a fixed 3000ms window and the pattern of results was comparable to those with the individually defined analysis windows. Within the response period, response durations were defined as the proportion of time in the analysis window that the animal spent looking towards the presenting audio speaker beyond 3SD of the baseline fixation period (Figure 3.3B). For analysis of the average looking-response to individual elements, the window was the time during which the element was presented with an adjustment for how long, on average, it took the animal to breach the 3SD criterion to look towards the start of the test sequence. The average horizontal eye deflections during the stimulus presentation were also analysed.

3.5.2. Results

Eye-tracking responses were analysed using an RM-ANOVA with two factors: 'monkey' (3 levels) and 'sequence condition' ('consistent' and 'violation'). The results confirmed those seen in the video-coding experiment. The animals made significantly longer looking-responses to the violation sequences than the consistent sequences (significant main effect of sequence condition: $F_{2,73} = 20.297$, $p < 0.001$; Figure 3.5A). Although individual animals differed in their looking times towards the presenting audio speaker (significant main effect of monkey, $F_{2,73} = 4.055$, $p < 0.05$), there was no interaction between sequence condition and monkey factors ($p = 1.0$), suggesting that while there may have been differences in responsiveness, all monkeys showed a preference for the violation sequences.

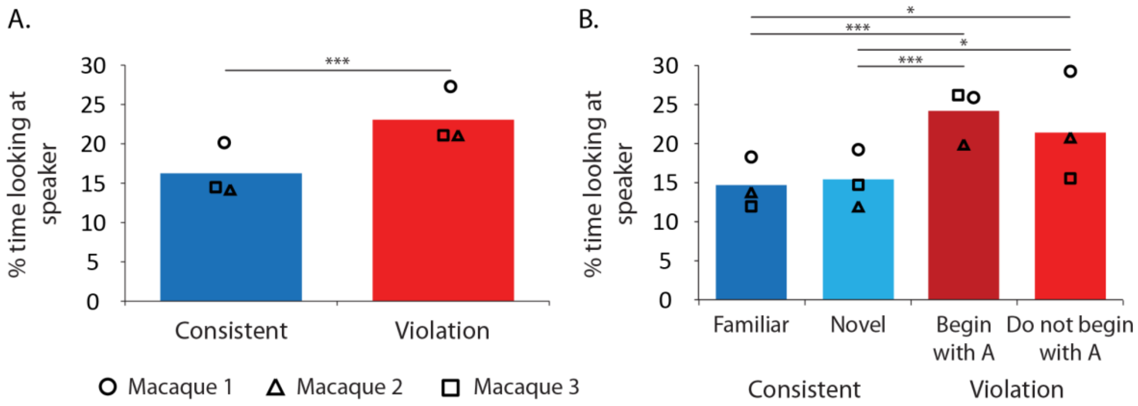


Figure 3.5 Group eye-tracking results. (A) Mean proportion (%) of looking-responses to the consistent and violation conditions. (B) Results for different subsets of the consistent and violation sequences. Shown are the results for the 'familiar' and 'novel' consistent test sequences and violation sequences that (like the consistent sequences) 'begin with A' or those that 'do not begin with A'. * = $p < 0.05$, *** = $p < 0.001$.

An ANOVA with four levels of the factor: ‘sequence condition’ (‘familiar’, ‘novel’, ‘begin with A’ and ‘do not begin with A’; Figure 3.1) showed a significant main effect for sequence condition ($F_{3,219} = 10.057$, $p < 0.001$; Figure 3.5B). Bonferroni comparisons showed significant differences were seen between: (1) ‘novel’ and ‘begin with A’ ($p < 0.001$); (2) ‘novel’ and ‘do not begin with A’ ($p = 0.01$); (3) ‘familiar’ and ‘begin with A’ ($p < 0.001$); and, (4) ‘familiar’ and ‘do not begin with A’ ($p = 0.01$; Figure 3.5). There was no significant difference between responses to ‘familiar’ and ‘novel’ consistent sequences ($p = 1.0$), nor any difference between responses to the violation sequences which ‘begin with A’ or ‘do not begin with A’ ($p = 1.0$). Again, there was no interaction between ‘sequence condition’ and ‘monkey’ factors ($p = 1.0$). These results replicate those of the video-coding experiment with increased statistical power, showing that macaques are able to discriminate consistent from violation sequences. Furthermore, these results suggest that the animals’ AGL abilities do not depend on sequence familiarity or only noticing violation sequences that do not begin with A.

As well as analysing the duration of the looking responses I calculated the mean horizontal eye position of each animal throughout the stimulus period in the direction of the speaker presenting the test sequence. An RM-ANOVA with two main factors: ‘monkey’ (3 levels) and ‘sequence condition’ (‘consistent’ and ‘violation’) supported the previous analyses showing a main effect of sequence condition ($F_{1,73} = 10.759$, $p = 0.002$). Although individual animals differed in the degree to which they looked towards the presenting audio speaker (significant main effect of monkey, $F_{2,73} = 3.666$, $p = 0.03$), there was no interaction between sequence condition and monkey factors ($p = 0.388$). This suggests that the animals showed a similar pattern of responses, although the magnitude of the effect may vary. A second RM-ANOVA with four levels of the factor: sequence condition (‘familiar’, ‘novel’, ‘begins with A’ and ‘does not begin with A’) again showed a significant main effect for sequence condition ($F_{3,219} = 3.60$, $p = 0.014$) and of monkey ($F_{2,73} = 3.666$, $p = 0.03$) and no significant interaction ($F_{6,219} = 1.958$, $p = 0.07$). Bonferroni comparisons showed significant differences only between: ‘novel’ consistent sequences and violation sequences which ‘do not begin with A’ ($p = 0.017$), all other contrasts were non-significant. While these effects are less robust than the previous analyses based on looking duration, which are not so easily biased by strong looking deflections in individual trials, the results are complementary. Both analyses suggest that the monkeys respond more strongly to sequences that violate the AG structure compared to those that are consistent with it, that this cannot be attributed to only familiarity or noticing initial violations, and that these results appear relatively consistent across animals.

To investigate this in more detail the more robust looking durations were analysed in individual animals (Figure 3.6). Significantly longer responses to violation compared to consistent sequences were observed in two monkeys ($t_{24} = 3.137$, $p = 0.004$; $t_{24} = 3.129$, $p = 0.005$) and the third was at the significance threshold ($t_{25} = 2.023$, $p = 0.05$). When all 4 sequence conditions were compared in three separate RM-ANOVAs all three monkeys showed a significant main effect of condition ($F_{3,72} = 3.715$, $p = 0.015$; $F_{3,72} = 4.745$, $p = 0.004$; $F_{3,75} = 5.08$, $p = 0.003$). Bonferroni corrected post-hoc tests are highlighted in Figure 3.6. No monkeys showed different responses to novel vs. familiar test sequences. Against predictions, monkey 3 shows stronger responses to violation sequences which 'begin with A' (i.e., are from this first element very much like the consistent sequences) compared to those which 'do not begin with A'. Thus, none of the monkeys preferentially responded only to sequences that violate the initial 'A' element. Due to reduced statistical power in individual animals, only one monkey showed significant, Bonferroni corrected, response differences between 'novel' sequences and violation sequences that 'begin with A' (Figure 3.6B). However, in combination these results suggest that all the macaques were sensitive to violations of the AG structure, and that none of the main effects can be attributed to either a familiarity effect or animals only responding to violations of the first element.

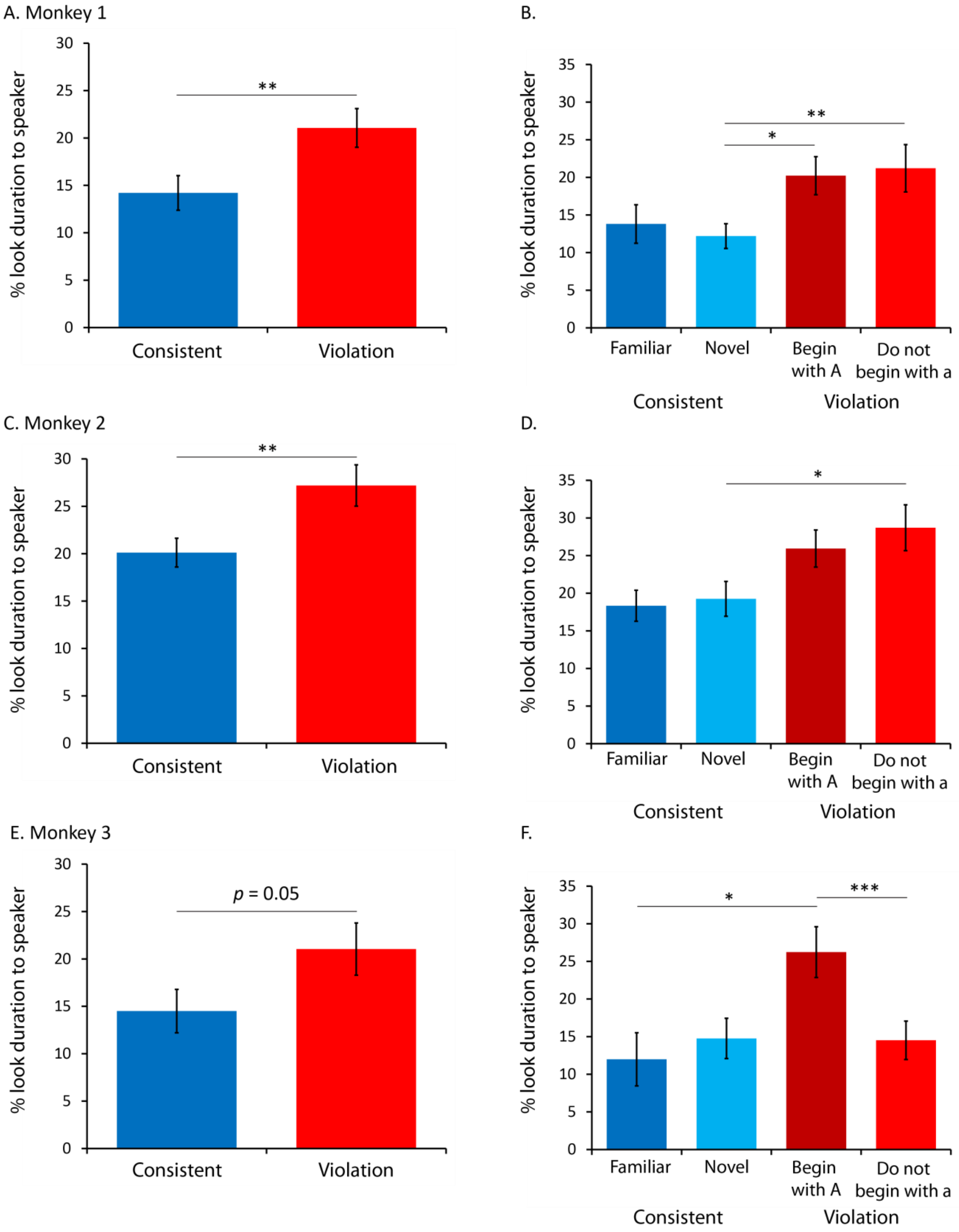


Figure 3.6 Eye-tracking results in individual monkeys. Eye-tracking effects for each monkey showing the proportion (% ±SEM) of looking responses to consistent and violation conditions (left panels) and to the four different stimulus conditions ('familiar' and 'novel' consistent sequences and violation sequences which 'begin with A' or 'do not begin with A'; right panels). * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

These results recapitulate the video-coding results and suggest that the macaques' abilities to discriminate consistent from violation responses do not depend on sequence familiarity or on rote memorisation during the habituation phase. Given that the monkeys seem to monitor the sequences for violations after the first element (they do not respond more strongly or for longer durations to violation sequences that 'do not begin with A'), does this mean that they can monitor for possible violations throughout the rest of the sequence? In particular, do they respond to violations beyond the 2nd position in the violation sequences, at which point the branching structure of the AG becomes more evident (Figure 3.1)?

To answer these questions and to better determine the extent of macaque AG learning abilities, eye movements in response to identical acoustical elements in either consistent or violation sequences were compared (Figure 3.7) with an RM-ANOVA with the factors of 'sequence condition' ('consistent' or 'violation'), 'element' ('A', 'C', 'D', 'F' or 'G') and 'monkey' (3 levels). Critically, the main effect of sequence condition ($F_{1,73} = 11.978$, $p = 0.001$; Fig. 4) did not interact with element ($p = 0.1$), nor was there a main effect of element itself ($p = 0.6$). Thus the stronger looks to violation sequences cannot be explained by the animals' responses to any individual element but only by the context in which the element occurred.

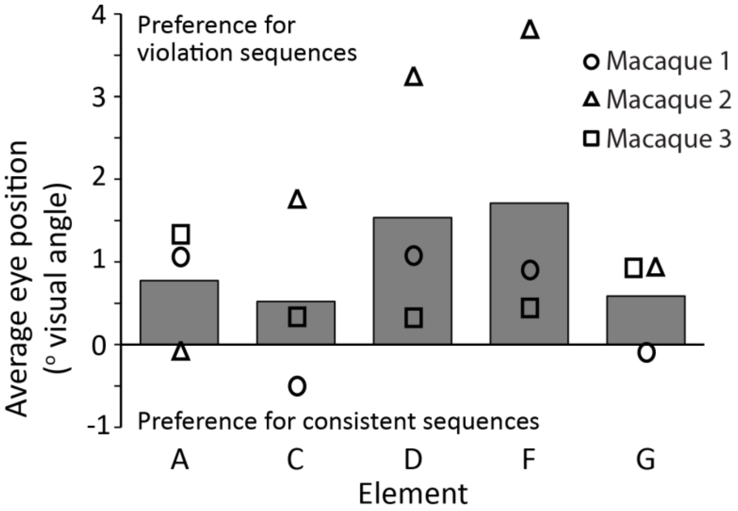


Figure 3.7 Contextual effects by AG element. Group (and individual) mean difference plot of responses to 'violation' – 'consistent' sequences towards the presenting speaker in response to each of the five stimulus elements (A, C, D, F and G). Positive differences reflect stronger looks to violation sequences than to consistent sequences.

To assess whether the macaques were sensitive to subtle, additional violations in later parts of the testing sequences, I analysed whether the macaques responded differently to two specific test sequences (*i* and *ii* in Figure 3.8A). These sequences begin identically, have their initial violation in the transition between the first and second positions, and contain the same elements in positions 3-5. However the elements in positions 3-5 have a different order, which generates an additional violation in sequence *ii* between the second and third element. Mean difference plots between sequence *i* and *ii* were generated for each monkey (Figure 3.8B-D). 95% confidence intervals were generated using a bootstrapping procedure as follows. Within the early part of the sequence, during the presentation of the first two elements, which is identical between the sequences, a data matrix of the eye-position throughout this period by the number of repeats of the two sequences was created. The sequence labels were then shuffled 1000 times to generate the null distribution of differences. By calculating the mean eye trace of these shuffled datasets, 5% and 95% confidence intervals (CIs) were calculated (Figure 3.8). Deviations of the difference eye trace below the 5% CI reflect responses is in favour of sequence *i*, with fewer violations, differences above the 95% CI would show a preferences for sequence *ii*. Lastly, for any significant deflection below the 5% or above the 95% CI, the area (representing both the time and magnitude of the deviation across the CI) that breached this significance threshold was calculated.

The results demonstrated that two of the macaques showed strong significant responses in favour of sequence *ii*, containing the additional violation, which resulted in an area above the significance threshold at least a factor of 3 greater than any such preference seen either for sequence *i* or during sequence positions 1-2 where the two sequences are identical. No difference could be observed between the sequences in macaque 3 (Figure 3.8D). These results suggest that a significant sensitivity to a subtle violation later in the sequences can be measured in a majority of the three animals studied.

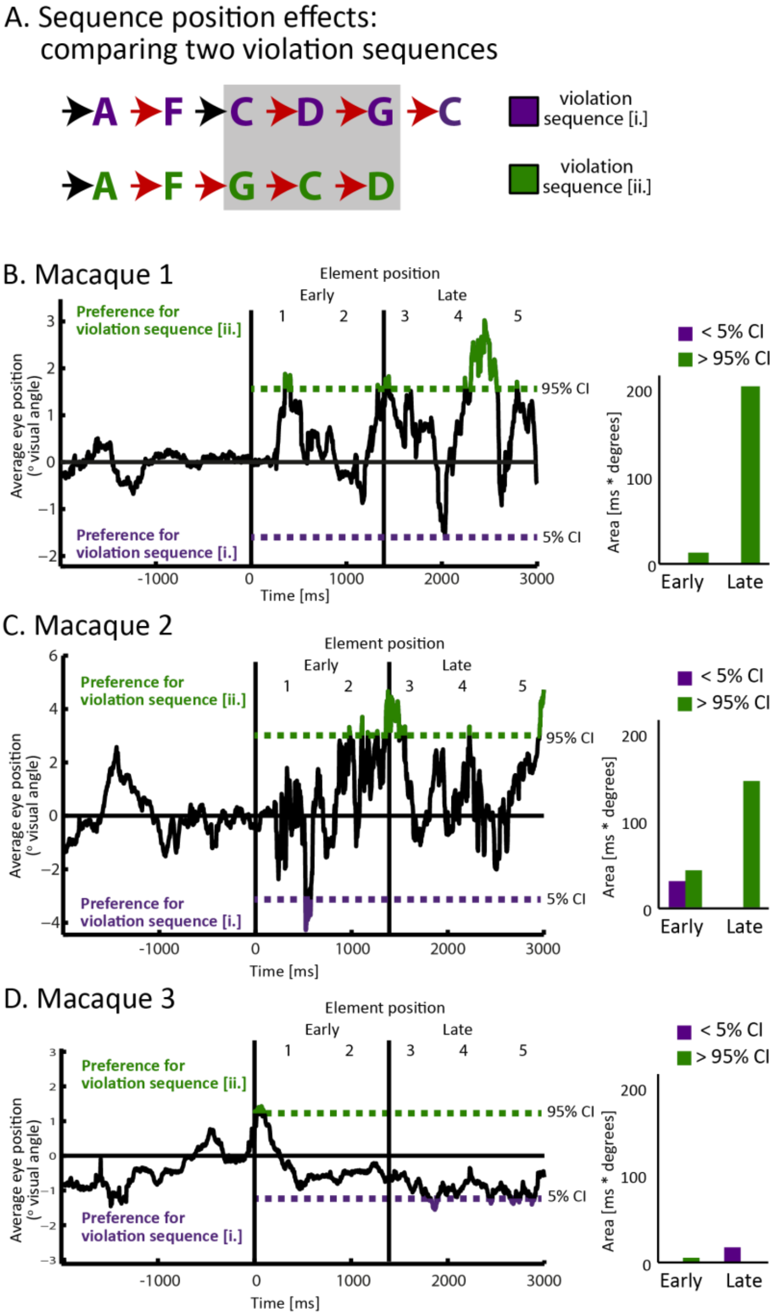


Figure 3.8 Eye-tracking sensitivity to violations at specific positions in the sequence. (A) Schematic plot of two of the violation sequences, identifying legal transitions (black arrows) and violations (red arrows). Violation sequence *ii* (green) contains one extra violation than sequence *i* (purple) between the second and third elements in the sequence. (B-D) Eye-tracking difference plots of preferences for sequence *ii* (positive numbers) or sequence *i* (negative numbers) including bootstrap determined confidence intervals (CI, dashed horizontal lines) based on the bootstrapped difference (1000 permutations) during element positions 1-2 which are identical between the two sequences. Also shown is the area above 95% or below 5% CI (bar plots on the right) where each animal made statistically significant looks in favour of either sequence. None of the 3 macaques showed a considerable preference for either sequence in the early part of the analysis (element positions 1-2). However, two of the macaques (B-C) showed at least a 3 factor increase in the area above the 95% CI in favour of sequence *ii* after the extra violation had occurred (later positions 3-5). (D) Macaque 3 does not show such a sensitivity.

A further analysis of the average look magnitude toward the presenting speaker during element positions 3-5 showed a significant preference for sequence *ii* in this period ($z = 1.994$, $p = 0.046$; Figure 3.9), or in analysis of the later positions 4-5 ($z = 2.173$, $p = 0.03$), for the data from macaques 1 and 2. An analysis of the data from all three animals (including macaque 3 that showed no sensitivity) recapitulates those reported in Figure 3.9, showing a trend for the analysis of positions 3-5 ($z = 1.677$; $p = 0.09$) and an effect for the later element positions 4-5 ($z = 1.943$; $p = 0.049$). Together these results suggest that the majority of macaques (2 out of 3) noticed violations throughout the AG sequences rather than just at the beginning. Taken together, these results suggest that Rhesus macaques recognise violations throughout the course of the sequences, and that this novel eye-tracking approach is sufficiently sensitive to measure these responses.

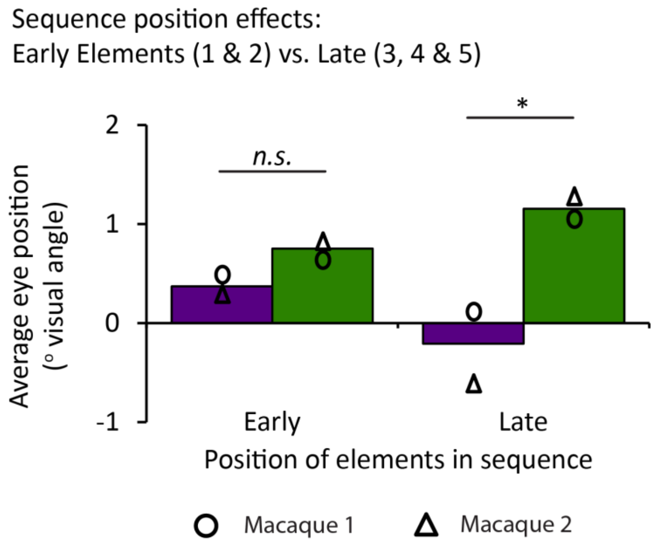


Figure 3.9 Eye-tracking responses to early and late parts of violation sequence ‘*i*’ and ‘*ii*’. In two animals (Monkey 1 and 2), significantly stronger eye-tracking responses to the sequence with additional violations are observed in the late, but not the early part of the sequence.

3.6. Discussion

The results reported here show that Rhesus macaques are sensitive to violations of a forward branching, non-deterministic AG, which has been insufficiently demonstrated in previous studies (see Chapters 1 & 2, Berwick *et al.*, 2011; Beckers *et al.*, 2012). The video-coding results demonstrate that, at the group level, macaques respond more frequently, more strongly and for longer durations to sequences which violate the artificial grammar. Furthermore, these results cannot be explained by trivial explanations such as rote memorisation or responding to novel stimuli, or by only responding to potentially highly salient violations (in the first position of the sequence). A complementary eye-tracking experiment recapitulated these results, including analyses in individual animals. Moreover, the additional accuracy and temporal specificity of the eye-tracking paradigm provides insights into responses to AG violations which would be impossible with more traditional methods. The results help to demonstrate that simple learning strategies appear to be insufficient to explain the macaques' responses, and that responses cannot be attributed to simple preferences for certain acoustical stimuli. Furthermore, analysis of two specific violation sequences demonstrate that at least two out of the three monkeys tested show preferences for the sequence with additional violations, late in the sequence. In summary, these results demonstrate that Rhesus macaques showed previously unreported sequence-structure learning capabilities. These findings support the hypothesis that the ability to evaluate non-deterministic, sequential structures may be subserved in certain primates, by generic, rather than language specific, processes.

3.6.1. Technical advances

One of the important implications of this work is the development of an objective method of measuring natural response behaviour in nonhuman primate AGL experiments. All previous research in nonhuman primates has relied upon video recording the animals' responses to test sequences for later offline analysis (Hauser *et al.*, 2001; Fitch & Hauser, 2004; Newport *et al.*, 2004; Saffran *et al.*, 2008; Hauser & Glynn, 2009). These methods contain an unavoidable level of subjectivity, which can only be overcome with an automated technique such as the one described here. The video-coding results reported here support the more objective eye-tracking results, both providing validation to this method and supporting the validity of the video-coding approach. Nevertheless, the eye-tracking technique provides not only valuable objectivity but also greater accuracy and temporal precision, presenting an improvement to traditional experimental designs.

3.6.2. Origins of primate AGL abilities

Human AGL abilities are thought to be supported by some of the same cognitive mechanisms and brain areas as natural language processing (Petersson *et al.*, 2004; Friederici *et al.*, 2006a). However, nonhuman primates clearly do not possess such linguistic abilities. Furthermore, unlike some other Old World monkeys (e.g. putty-nosed monkeys, Arnold & Zuberbuhler, 2006), there is no evidence macaques meaningfully combine their calls or that their vocalisations contain any discernible structure. Therefore, it may be necessary to distinguish between vocal production and auditory learning processes since these capacities seem to be subserved by different neurobiological pathways and mechanisms (Jarvis, 2004; Petkov & Jarvis, 2012). Rather than vocal communication, the sequence-structure learning abilities that support this AGL task may relate to other capabilities. The natural function of these abilities in nonhuman primates is remains an empirical question. It is possible that nonhuman primates' capabilities to learn aspects of AG structures may arise from their abilities to evaluate patterns in sensory input (or the structure of social interactions: Bergman *et al.*, 2003; like the movement patterns of others: Schmitt, 2010). However, behavioural experiments in a single nonhuman species are likely to be insufficient to address this question. Comparative, cross species analyses may allow AGL capabilities in a number of species to be analytically compared to other cognitive or behavioural abilities to attempt to assess the likely natural functions of these abilities. Alternatively, neuroimaging experiments might reveal that AGL tasks produce activity in specific brain areas with well-understood functions, (e.g. motor areas), suggesting that it is these abilities that are recruited for artificial grammar learning. While the ultimate roots of the sequence-structure learning abilities required for AGL tasks are currently unclear, experiments such as this one at least provide a rigorous way to investigate the abilities the animals possess, and therefore may represent the basis for further research into these questions.

3.6.3. Remaining questions and future directions

The results of this experiment provide the first robust evidence that nonhuman primates are able to learn an AG structure of this level of complexity. However, understanding the abilities of a single species to learn one type of AG structure has limited benefits in understanding the evolution of how the abilities supporting this learning may have evolved. Would other nonhuman animals be able to learn this AG under similar conditions? Would New World monkeys, which are more distantly related to humans but tend to be more vocal than Old World monkeys (DeVore, 1963; Fooden, 2000), perform better, worse, or comparably to Rhesus macaques? In Chapter 4, I will present data from both common marmosets, a species of New World monkey, and human participants, assessing their abilities to learn the same AG structure. The results of these experiments will facilitate better mapping of the evolutionary origins of these sequence-structure learning capabilities and how they relate to the language faculty in humans.

A second question raised by these results is, how do Rhesus macaques process these sequences, and does it relate in any way to how the human brain processes language? It is possible that the relatively simple (compared to human language) sequencing abilities required for these tasks are subserved by generic processes, shared between humans and other extant primate species. This would allow humans and macaques to process the AG sequences, and detect violations of the AG structure, in comparable ways. However, it is also possible that while humans possess a specialised network of brain areas, evolved to process language, which makes such AGL relatively trivial, nonhuman primates may employ completely different cognitive mechanisms supported by different brain areas to produce comparable results. In Chapter 5 I will present comparative fMRI data from Rhesus macaques and human participants, investigating the brain regions involved in processing this AG structure to explore similarities and differences between how AGL in the human and nonhuman primate brain.

3.7. Conclusion

These two experiments revealed a previously unknown level of complexity in Rhesus macaque AGL capabilities. Such behavioural results provide an important initial foundation required to begin asking both how the human language faculty may have evolved and how useful these nonhuman primates may be as an animal model system. However, to address either of these questions, it is necessary to consider how other species AGL may relate to those of the Rhesus macaque, and how AGL may be processed in the brains of these animals. In the following chapters I will attempt to address these questions, with the ultimate goal of furthering our understanding of how the human language faculty may have evolved, and whether nonhuman primates may one day help to reveal the neuronal mechanisms underlying aspects of human language.

Chapter 4: Comparative AGL in Macaques, Marmosets and Humans

Parts of this chapter have been published as:

B. Wilson, H. Slater, Y. Kikuchi, A. E. Milne, W. D. Marslen-Wilson, K. Smith & C. I. Petkov (2012). Auditory artificial grammar learning in macaque and marmoset monkeys. *The Journal of Neuroscience,* **33(48)**, 18825-18835.

All work presented in this chapter was primarily conducted by the author, except where explicitly stated. Co-authors on this paper provided valuable comments and discussion, and H. Slater and A. Milne provided additional ratings of the marmosets' responses. The majority of the experimental design, the running of the experiments, data analysis and writing were conducted by the first author.

4.1. Abstract

Artificial Grammar Learning (AGL) paradigms present an opportunity to comparatively study how non-linguistic species may learn structures that emulate specific features of natural language. While a number of studies, including those reported in the previous chapter, have demonstrated AGL in nonhuman animals, no study has yet directly compared the abilities of different species to learn the same AG structure. Such comparisons are necessary to place these AGL capabilities within a comparative framework, to potentially provide insights into how language may have evolved in humans and to aid in the development of animal model systems for the investigation of language related capabilities. Following previous experiments in Old World monkeys; common marmosets (a species of New World Monkey) and human participants were tested using the same AG structure. In a two-alternative, forced-choice experiment, human participants quickly learned to differentiate between sequences of nonsense words that were consistent with the AG and those that violated the AG structure. In an experiment identical to the previous video-coding experiment in macaques (see Chapter 3), common marmosets also showed response preferences for sequences that violated the AG structure. However, unlike the macaques, these responses appeared to be driven by a reliance on simpler strategies, including primarily noticing violations in the first sequence position or the novelty of the test sequences. These results provide some support for the hypothesis that species more closely related to humans may possess more complex AGL capabilities, suggesting a potential evolutionary gradient for sequence-structure learning complexity. These results are discussed with respect to how comparative analyses might inform us on the topic of language evolution, and the potential suitability of each species as an animal model system for aspects of language processing.

4.2. Introduction

The experiments in the previous chapter demonstrated that Rhesus macaques are able to learn the structure of a forward-branching, non-deterministic artificial grammar, and that these results cannot be easily attributed to simple cues. Evidence for this complexity of AGL capabilities in macaques demonstrates that these abilities are not uniquely human, and therefore are unlikely to be specific to language. Furthermore, behavioural evidence that an animal possesses comparable abilities to humans suggests that it might make a valuable animal model system to explore such abilities at a level impossible in humans. However, experimentation in a single species is limited in the extent to which it can inform us about issues regarding language evolution. While the presence of this ability in a single nonhuman primate suggests that the capacity may be evolutionarily conserved from a common ancestor, it is possible that both species may have independently developed such a capability. Many abilities and traits have evolved independently in different lineages by convergent evolution. For example, vocal learning in songbirds (and other, distantly related species) is known to have evolved independently multiple times (Jarvis, 2004; Feenders *et al.*, 2008; Petkov & Jarvis, 2012). Therefore, the co-occurrence of an ability in both humans and a single nonhuman primate may not be sufficient to generalise the capacity to other nonhuman animals. Alternatively, it is possible that such sequence-structure learning abilities may be very common in nonhuman animals. If this were to prove to be the case, evidence in Rhesus macaques alone might tell us relatively little about when the ability may have evolved. Testing in additional species presents the opportunity to better understand the prevalence of these abilities in nonhuman primates and therefore to inform us about language evolution. Furthermore, the study of a wider range of species might provide additional candidates to act as animal models for aspects of language processing. While a single species might represent an adequate animal model, additional factors, such as similarities in neurobiology, ease of housing and breeding or ethical issues, might make other species more appropriate.

The sequence-structure learning capabilities required in AGL paradigms require relatively complex abilities. Therefore, it is likely that rather than a binary distinction between being either present or absent in any given species, these abilities might exist to varying extents in a wide range of animals. For example, previous AGL experiments have shown that cotton-top tamarins (a New World monkey) are able to learn simple AG structures but not more complex ones (Fitch & Hauser, 2004), and that their abilities might be based on relatively simple learning strategies (e.g. familiarity of the test sequences, Saffran *et al.*, 2008). In contrast, the AG used here is more complex than those used by Fitch and Hauser (2004, see Chapter 2), and, unlike the tamarins tested with a similar grammar (Saffran *et al.*, 2008), the responses of the Rhesus macaques could not be attributed only to simple strategies (see Chapter 3). Therefore, might the study of additional species demonstrate that more closely related animals have more complex AGL abilities? While the parameter space proposed in Chapter 2 presents an objective way to compare AGL abilities across different AG structures, direct comparisons between species require the same animals to be tested using the same AG structure and methods.

Artificial grammar learning experiments were initially designed to investigate how human participants learned simple structures emulating features of natural language (Reber, 1967). Indeed, neuroimaging evidence suggests that some of the same areas involved in processing language are also involved in artificial grammar learning (Petersson *et al.*, 2004; Friederici *et al.*, 2006a). However, language in humans is undoubtedly a unique evolutionary specialisation, absent in nonhuman animals. Therefore, what natural abilities in nonhuman animals are these AGL tasks tapping into? As has previously been discussed (see Chapter 3), behavioural evidence in a single species is unlikely to identify the natural abilities which underpin these sequence-structure learning abilities. However, the study of additional species might help to clarify this. For example, if AGL abilities were based on a species' vocal production capabilities, we might expect those species which produce more complex or frequent vocalisations to also perform better on AGL tasks, possibly independent of the species' phylogenetic relationship to humans. To rigorously conduct such analyses it would be necessary to collect comparative data on the AGL capabilities, as well as other potentially related behavioural or physiological data, from a wide range of species. While a research project of this magnitude falls outside the scope of this thesis, the study of two nonhuman primates provides a number of important benefits compared to experiments in a single species.

Data in a second species is the initial step in building up a large enough dataset to perform the type of phylogenetic analyses required to answer complex questions about the evolution of language. In the shorter term, evidence of AGL in a second primate species could help to clarify whether, or to what extent, these sequence-structure learning abilities might be common to other nonhuman primate species. Such data also facilitates more accurate estimations of when these abilities may have evolved. Finally, evidence of AGL in a second species presents us with another possible animal model system in which aspects of human language learning abilities might be investigated.

In this chapter, I tested how the forward-branching, non-deterministic AG structure might be learned by common marmosets. Marmosets are a New World monkey that shared a last common ancestor (LCA) with humans approximately 45 million years ago (MYA). Rhesus macaques are an Old World monkey, and are more closely related to humans, (LCA approximately 30 MYA, Steiper & Young, 2006). Common marmosets are relatively closely related to cotton-top tamarins, another species of New World monkey, which have been the subject of many of the previous nonhuman primate AGL studies (Hauser *et al.*, 2001; Fitch & Hauser, 2004; Newport *et al.*, 2004; Saffran *et al.*, 2008). However, the current experiment represents the first attempt to directly compare Old and New World monkeys using the same AGL paradigm. Furthermore, marmosets are a commonly used animal model system in neuroscience (Okano *et al.*, 2012). Therefore, the assessment of their AGL abilities presents the possibility to consider these animals, as well as Rhesus macaques, as animal models in which some language related abilities might be studied at the cell or molecular level. Eye-tracking experiments, as described in macaques in Chapter 3, were not feasible in the marmosets due to their small size. However, the results of the macaque eye-tracking and video-coding experiments were complementary; therefore, the video-coding experiments appear to provide a valid measure of AGL in nonhuman primates (also, see Discussion).

In this chapter, I will also present data on the abilities of human participants to learn the same AG structure. Previous evidence suggests that human adults, and even infants, are able to learn non-deterministic AG structures (Reber, 1967; Saffran, 2002; Petersson *et al.*, 2004; Saffran *et al.*, 2008). While the evidence that humans can learn AGs of this type is uncontroversial, testing participants using identical stimuli presents a number of advantages. Beyond providing behavioural confirmation of these AGL abilities in humans, the data from this experiment can be directly compared to those collected in other species, including analyses based on the familiarity of sequences or the extent to which responses can be attributed to simple cues. Furthermore, neuroimaging experiments require behavioural confirmation of AGL in individual participants (see Chapter 5). The human experiment used identical stimuli sequences to the nonhuman primate experiments and similar habituation and testing phases. However, natural response experiments, such as those performed in nonhuman primates, are difficult to conduct in human participants. Therefore, the testing phase of this experiment required participants to make explicit responses (see Methods and Discussion). While the testing measures differed between species, this approach was necessary to gather comparable data across species as different as humans and monkeys.

4.3. Rhesus macaque summary

To facilitate comparisons across species this section will briefly review the key results of the Rhesus macaque video-coding experiment. Following habituation to sequences of nonsense words consistent with the AG structure, the macaques were tested with consistent sequences (including 'familiar' and 'novel') and sequences that violated the AG structure (which either 'begin with A' or 'do not begin with A'). The animals' responses were videotaped and coded by three independent raters for the proportion of trials on which the animals responded and the duration and strength of these responses. Significant differences between consistent and violation sequences were identified for all response measures (Figure 3.2, reproduced as Figure 4.1). Furthermore, Bonferroni corrected post-hoc tests revealed significant differences between the key contrast of 'novel' consistent sequences and violation sequences which 'begin with A' (which cannot be solved by either familiarity of recognising the violations in the first element) for both the proportion of responsive trials and the strength of responses (Figure 4.1B and F). More liberal post-hoc comparisons revealed, for every response measure, significant differences between consistent and violation conditions (i.e. 'familiar' vs. 'begins with A'; 'familiar' vs. 'does not being with A'; 'novel vs. 'begins with A'; 'novel vs. 'does not being with A') but never within a condition (i.e. 'familiar' vs. 'novel'; 'begins with A' vs. 'does not begin with A'). These results suggest that the macaques learned the AG structure, and that their responses cannot be attributed to either the familiarity of the sequences or responding only to simple rule violations, such as noticing violations in the first sequence position. For more details, see Section 3.4, and Figure 4.1.

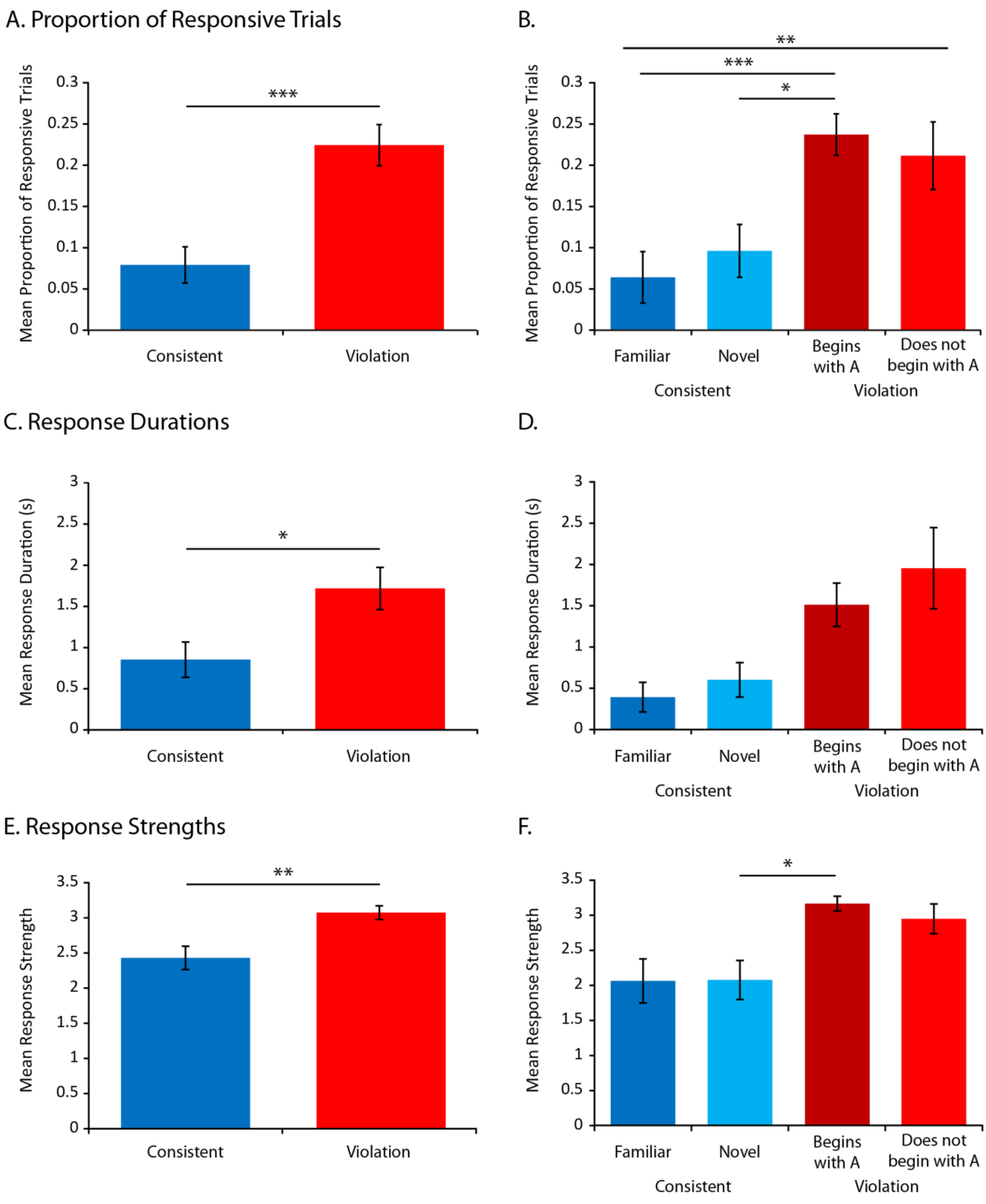


Figure 4.1 Reproduction of Figure 3.2: Video-coding results in Rhesus macaques. (A-B) Mean proportion (±SEM) of trials on which macaques made unambiguous looking-responses as evaluated by a majority of raters (see Methods). The left panels indicate responses to consistent and violation sequences, right panels display results to specific subsets of stimulus sequences. (C-D) Mean response duration of responses (±SEM) to consistent and violation sequences (C) and subsets of sequences (D). (E-F) Mean responses strengths based on coding from three raters (±SEM) to consistent and violation sequences (E) and subsets of sequences (F). Significance levels are shown for all main effects (paired sample *t*-tests; left panels) and Bonferroni post-hoc tests (right panels); * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

4.4. Common marmosets experiment

4.4.1. Methods

To facilitate cross species comparisons the methods for this experiment were kept as similar as possible to those of the Rhesus macaque experiment. However, only 4 marmosets were available for testing, therefore each monkey was tested on 4 different occasions separated by at least one week (see below).

Participants

Four common marmosets (*Callithrix jacchus*) participated in this experiment. The marmosets were housed in two pairs within a single colony, and were individually housed for testing. All animal work and procedures were approved by the U.K. Home Office and abide by the Animal Scientific Procedures Act 1986 on the care and use of animals in research.

To assess whether any potential differences observed between the species could be attributed to different levels of auditory perception the power spectrums of the nonsense word stimuli were compared to human, macaque and marmoset audiograms (Figure 4.2). I confirmed that the power spectrum density of the nonsense word stimuli was well within the audible range of all of the species (i.e., at least 30dB above their hearing threshold in the range of ~100-5000Hz). Therefore, any differences between the species are unlikely to result from differences in auditory perception.

Stimuli and procedure

The stimuli, habituation and testing procedures and video-coding analysis used in this experiment were identical to those used in the macaque video-coding experiment (see Section 3.4.1). To obtain sufficient results with the 4 marmosets that were available for testing, the animals were tested on 4 separate occasions, with at least 1 week between the testing sessions. Additional analyses confirmed that the results could not be explained by individual differences or any cumulative learning effects (see Results).

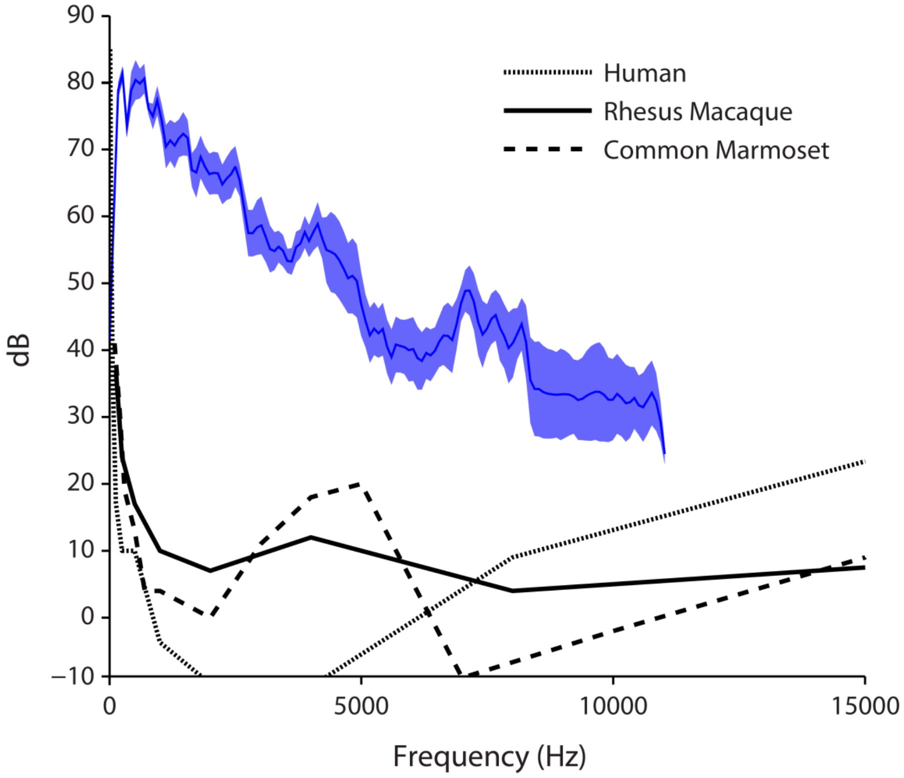


Figure 4.2 Nonsense word power spectra and primate audiograms. Mean (± SEM) acoustic power spectrum of the five nonsense word stimuli (in dB re 20 μN/m^2, blue line) . The black lines indicate the auditory threshold above which sounds at that frequency and intensity are audible for humans (dotted line), macaques (solid line) and marmosets (dashed line). The nonsense word stimuli fall well within the audible range of all three species, and there are no gross differences in frequency sensitivity between the species. Human audiogram data from: (Jackson *et al.*, 1999); macaque audiogram data summarised from: (Pfingst *et al.*, 1975; Pfingst *et al.*, 1978; Lonsbury-Martin & Martin, 1981; Bennett *et al.*, 1983); marmoset audiogram data from: (Seiden, 1958)

Data analysis

As with the macaque data, for all analyses only trials on which the majority of the raters (2 out of 3) agreed that an unambiguous response was made (strength of response ≥4, on the scale of 1-5) were used. The proportions of trials on which the marmosets responded, as well as the duration and reported strength of these responses were analysed. Due to limited availability of animals, each of the 4 available marmosets was tested on four separate occasions. In order to ensure that differences in responses, on any of the measures collected, could only be attributed to the animals detecting violations of the AG, rather than to differences between animals or between sessions, the data were analysed using linear mixed-effects models (Bates, *et al.*, 2012) using the R software package, as follows. Linear mixed effects models were constructed to test the extent to which each response measure (response duration, response strength, or response proportion) was predicted by 'sequence condition' ('consistent' or 'violation', fixed effect), controlling for the random factors: 'marmoset' (4 levels) and 'session' (4 levels). All analyses used random slope models (Field, 2009; Winter, 2013) because it is possible that beyond producing different levels of responses in general (as would be modelled by a random intercept model, Field, 2009; Winter, 2013), different marmosets or testing sessions might be associated with more or less responses in one particular condition (i.e. consistent or violation). To test the significance of the effects produced by each of these models, it was necessary to test an equivalent null model for each, which measured the extent to which the each response measure (response duration, response strength, or response proportion) could be predicted only by 'marmoset' and 'session' (both random factors). These two models were then compared using a likelihood ratio test (Winter, 2013), which tests the null hypothesis that the effect size obtained by the full model could also have been produced by the null model. Therefore, this analysis assesses the contribution 'sequence condition', which was the only difference between the models (see Results).

Inter-rater reliability

Three raters coded all 32 trials of all of the videos. Inter-rater reliability was calculated pairwise between the three raters. The raters had exact agreement on the strength of the response (on the five point scale) on 49.8% of the trials and were within one response point from each other on 80% of the trials. Cohen's Kappa (Landis & Koch, 1977) revealed "fair" to "moderate" mean agreement between raters, $K = 0.39$. Of all of the available trials the marmosets were rated as unambiguously responding on 22.8% of all of the recorded trials by a majority of raters resulting in 60 grammatical and 57 ungrammatical response trials (total of 117) used for analysis. These numbers, in comparison to those of macaques (see Section 3.4), indicate that the marmoset data were not statistically underpowered in relation to those that were available for analysis from the macaques.

4.4.2. Results

The frequency of marmosets' responses (response proportions) were analysed with linear mixed-effects models with the factors 'sequence condition' ('consistent' or 'violation', fixed effect), 'marmoset' (4 levels, random factor) and 'session' (4 levels, random factor), to assess the extent to which the animals' responses could be attributed to the grammaticality of the sequences, rather than individual differences between marmosets or differences between sessions. While the Rhesus macaques responded significantly more frequently to violation compared to consistent sequences, no such effect was observed in the marmosets (likelihood ratio test between experimental model, containing the 'sequence condition', 'marmoset' and 'session' factors, and the null model, containing only the 'marmoset' and 'session' factors: $\chi^2_1 = 0.0764$, $p = 0.782$, Methods, and Figure 4.3A). Similarly, additional analyses including 4 levels for the factor: 'sequence condition' (levels: 'familiar', 'novel', 'begins with A' and 'does not begin with A') revealed no differences between the experimental model and the null model ($\chi^2_1 = 0.02$, $p = 0.887$, Figure 4.3B). Post-hoc analyses were performed by testing models comparing individual pairs of sequence conditions (e.g., 'familiar' vs. 'novel') and using Bonferroni corrections for multiple comparisons. These analyses revealed no differences between any of the sub-conditions $p > 0.8$ in all cases). Interestingly, comparisons of the proportion of trials on which the two species responded suggested that the marmosets responded more frequently than the macaques, suggesting that while the marmosets might be more responsive than the macaques, these responses may be relatively indiscriminate with regards to the condition of the sequences (consistent or violation). Furthermore, because the marmosets were the more responsive species, more data contributed to the final analyses, therefore any results showing stronger effects in the macaques cannot be attributed to greater statistical power.

Although the marmosets appeared to respond equally frequently to the different sequences, responses to violation sequences had significantly longer durations than to the consistent sequences (responses to violation sequences were 0.7 seconds ± 0.34 (standard errors) longer than to consistent sequences, likelihood ratio test between linear mixed-effects models: $\chi^2_1 = 4.01$, $p = 0.045$, Figure 4.3C). Further analyses considering the four different conditions also showed a main effect of response duration ($\chi^2_1 = 5.08$, $p = 0.024$,

Figure 4.3D). Post-hoc analyses, with Bonferroni corrections, as above, revealed no significant differences between any sequence conditions ($p > 0.27$ in all cases). Furthermore, less conservative LSD multiple comparisons corrections only revealed a significant difference between 'familiar' sequences and violation sequences that 'do not begin with A' ($p = 0.045$), but no other contrasts reached even this very liberal significance threshold. Therefore, even using these very liberal corrections for multiple comparisons, no significant difference between the 'novel' sequences and those that 'begin with A' was identified. Thus it is difficult to conclude that the differences in marmosets' responses between consistent and violation sequences can be attributed to anything more than simple strategies, such as noticing only violations in the initial element or responding to novel, previously unheard sequences.

Finally, no differences in the ratings of the strength of the responses were observed between the consistent and violation sequences ($\chi^2_1 = 0.637$, $p = 0.43$, Figure 4.3E) or between the four levels of sequence condition ($\chi^2_1 = 0.583$, $p = 0.45$, Figure 4.3F) and post-hoc analyses, using either Bonferroni or LSD corrections, revealed no significant differences between conditions.

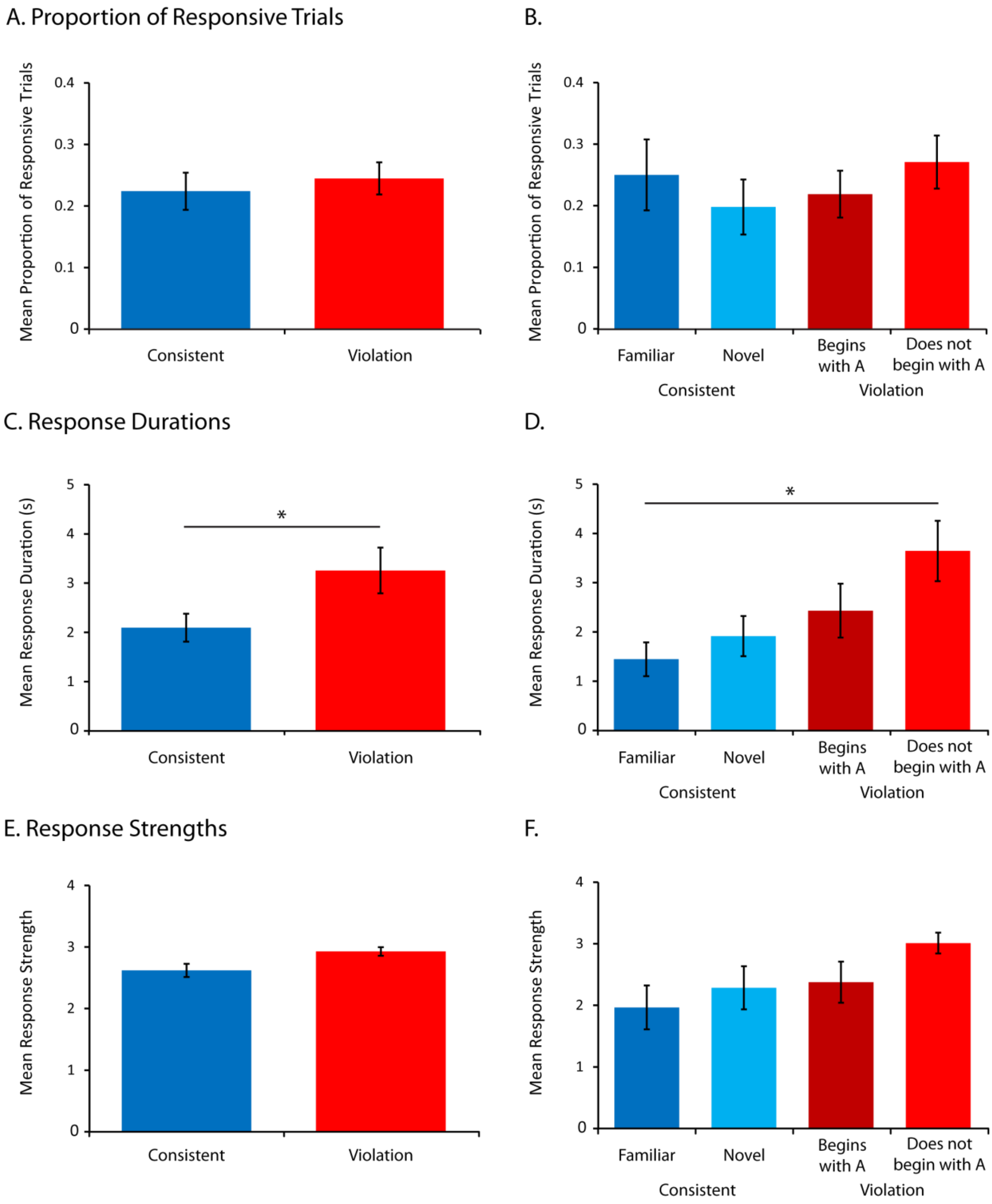


Figure 4.3 Video-coding experiment results in common marmosets. (A) Mean proportion (±SEM) of trials on which marmosets made unambiguous looking-responses as evaluated by a majority of raters in responses to consistent and violation sequences and (B) responses to ‘familiar’ and ‘novel’ consistent sequences and violation sequences that ‘begin with A’ and ‘do not begin with A’. (C-D) Mean response duration (±SEM) of responses. (E-F) Mean responses strengths (±SEM) based on coding from three raters. Significance levels are shown for all main effects (left panels) and post-hoc analyses (right panels); * = $p <$ 0.05.

4.4.3. Marmoset conclusion

These results suggest that in terms of frequency of responses, the common marmosets were relatively indiscriminate to consistent and violation sequences. Unlike Rhesus macaques, which responded significantly more frequently to violation sequences than consistent ones, no such differences were observed in the marmosets. Similarly, no differences between the consistent and violation sequences were seen in the strengths of the marmoset responses. However, the marmosets' orienting responses persisted for significantly longer durations following the presentation of violation than consistent sequences. This suggests that marmosets may initially attend to the presentation of any auditory stimulus, regardless of the sequence presented, but that they more quickly lose interest in the consistent sequences, while the violation sequences maintain their attention for longer. These increased looking durations, which were also evident in the macaques, suggest that marmosets are sensitive to some features of the AG structure.

The macaques responded similarly, in terms of both response frequency and duration, to the 'familiar' and 'novel' consistent test sequences, and to the violation sequences that 'begin with A' and that 'do not begin with A', however significant differences were apparent between the 'novel' consistent sequences and the violation sequences that, like the consistent sequences, 'begin with A'. Therefore the macaques' responses cannot easily be attributed to simple cues such as novelty or only detecting violations early in the sequences. By contrast, the common marmosets showed no such discrimination between these two key stimulus conditions, making it impossible to conclude that they did not rely primarily on these simpler strategies, or that they demonstrated meaningful learning of the AG structure. While discriminating between sequences based on such simple cues represents an early step in sequence-structure learning capabilities, they clearly do not represent a full understanding of the AG structure, or the level of AGL observed in Rhesus macaques.

4.5. Human experiment

4.5.1. Methods

The AG structure and stimuli sequences used to test humans were identical to those used in previous experiments. Human participants are difficult to test with implicit natural response paradigms such as those used in nonhuman primates. Adult humans, familiar with technology like audio speakers, do not typically produce strong dishabituation responses to repeated presentations of auditory stimuli. Therefore, it is very difficult to detect any differences between conditions that would demonstrate artificial grammar learning. In this experiment, humans, like the nonhuman primates, first took part in the implicit habituation phase (although humans required a shorter period of exposure to learn the AG structure, e.g. Saffran, 2002; Fitch & Hauser, 2004). However, in the testing phase the participants were asked to make explicit responses in a two-alternative, forced-choice experiment (see Discussion). Although participants were informed that there was a pattern to the sequences of nonsense words, they were given no information about the structure of the AG, and were given no feedback during the experiment. While this procedure represents a notable deviation from the methods of the primate experiments, this experiment provided a valuable data about human participants' abilities to learn the same AG structure the monkeys were tested with, allowing some comparison between the species.

Participants

Twelve adult participants (age range 19 to 34 years, mean age 23; 5 male, 7 female) were recruited through the Newcastle University Institute of Neuroscience participation scheme and provided informed consent to participate. All participants were native English speakers, had normal hearing and normal or corrected vision. No participants had any language or comprehension disorders. The ethics of this experiment were approved by the local ethics committee at Newcastle University.

Stimuli

The stimuli were identical to those used in previous experiments (see Section 3.4.1).

Procedure

Participants were seated one meter in front of a computer monitor and two audio speakers. All stimuli throughout the experiment were presented from both speakers. The experiment was run using custom Matlab scripts through Psychophysics toolbox. To generate sufficient data, and to assess ensure that the participants remained familiarised to the AG structure, testing phases were interspersed with additional habituation (re-familiarisation) phases. In total each participant took part in three testing phases, each preceded by a habituation or re-familiarisation phase.

Habituation phase

During the initial habituation phase, the participants were asked to listen to the habituation sequences (see Section 3.4.1) for 5 minutes, as a number of studies have demonstrated that adult and infant human participants require less habituation in order to learn an AG structure than nonhuman primates do, (Fitch & Hauser, 2004; Newport & Aslin, 2004; Newport *et al.*, 2004; Saffran *et al.*, 2008). Subsequent re-familiarisation phases presented the same habituation sequences, in a randomised order, for 3 minutes. The stimuli sequences were presented in an identical manner to previous, nonhuman primate experiments (rate of 9 sequences/min; inter-sequence interval = 4 sec). There was no task during this phase of the experiment, participants were only asked to sit quietly and listen to the sequences.

Testing phase

In the testing phase participants were presented with each of the test sequences 4 times in a pseudorandom order, for a total of 32 trials. During the stimulus sequence presentation a blue fixation spot was displayed on the screen. Following the presentation of each sequence, the fixation spot changed from blue to yellow, indicating that the participant should respond either that the sequence "followed the pattern" ('consistent') or "did not follow the pattern" ('violation'). The responses and response times were recorded. The participants were asked to respond as quickly and accurately as possible, but there was no time limit on responses. Following the participant's response, the next trial began after an inter-trial interval of 2 seconds.

Data analysis

To allow easier comparison to the nonhuman primate results, rather than plotting the mean percentage of correct responses, data are plotted as the proportion of trials to which the participants made the 'violation' response. Therefore for consistent conditions responses below the 50% chance level indicate good performance, and for violation conditions responses above the chance level indicate good performance. This allows clearer comparisons between the human and nonhuman primate data, as it facilitates not only comparisons of responses to the chance level, but also, as in the monkeys, between consistent and violation test conditions.

4.5.2. Results

The participants accurately identified consistent and violation sequences significantly better than chance levels (one-sample *t*-tests, consistent: $t_{11} = 8.193, p < 0.001$; violation: $t_{11} = 5.177, p < 0.001$; Figure 4.4A) and a significant difference was observed between conditions (paired sample *t*-test, $t_{11} = 7.681, p < 0.001$). However, when responses were analysed based on the four experimental conditions performance better than chance was observed in only three of the conditions (one sample *t*-tests, 'familiar': $t_{11} = 7.731, p <$ 0.001; 'novel': $t_{11} = 8.164, p < 0.001$; 'does not begin with A': $t_{11} = 7.330, p < 0.001$; Figure 4.4B) while violation sequences that 'begin with A' were not identified above chance levels ('begins with A': $t_{11} = 1.499$, $p = 0.162$). An RM-ANOVA revealed a main effect of 'sequence condition' ($F_{3,33} = 41.852$, $p < 0.001$). Bonferroni corrected post-hoc tests identified significant difference between all consistent and violation conditions (i.e., 'familiar' vs. 'begins with A', $p = 0.003$; 'familiar' vs. 'does not begin with A', $p < 0.001$; 'novel vs. 'begins with A', $p = 0.001$; 'novel vs. 'does not begin with A', $p < 0.001$; Figure 4.4B). No significant differences were observed between the two consistent conditions ('familiar' vs. 'novel', $p = 1.0$), suggesting the novelty of the test sequences played very little role in the participants decisions. However, participants correctly classified sequences that 'do not begin with A' as violation sequences significantly more accurately than sequences that 'begin with A' ($p = 0.025$), suggesting that participants more easily recognised sequences that contained violations in the first sequence position.

To investigate how participants' performance might improve over multiple testing runs additional analyses were conducted. Responses to consistent and violation sequences were separated by testing run (1-3). Responses to all testing runs were significantly better than chance (one-sample *t*-tests, $p < 0.01$ in all cases; Figure 4.4C). An RM-ANOVA with the factors 'sequence condition': (2 levels) and 'run' (3 levels) was performed, showed a main effect of sequence condition ($F_{1,22} = 59.002$, $p < 0.001$) but no main effect of run ($F_{2,22} = 0.823$, $p < 0.452$) and a non-significant trend towards an interaction between condition and run ($F_{2,22} = 2.858$, $p = 0.079$). These results show that participants' performance was high throughout the experiment, but the lack of a significant interaction suggests that performance did not significantly improve over testing runs. When the four different conditions were considered, the same pattern of results was observed. The consistent 'familiar' and 'novel' sequences and the violation sequences that 'do not begin with A' were classified better than chance on every run ($p < 0.001$ in all cases, Figure 4.4D). However the violation sequences that 'begin with A' were never classified significantly above chance (Run 1: $t_{11} = 0.194$, $p = 0.849$; Run 2: $t_{11} = 1.643$, $p = 0.129$; Run 3: $t_{11} = 19.58$, $p = 0.076$). An RM-ANOVA with the factors: 'sequence condition' (4 levels) and run (3 levels) revealed a main effect of sequence condition ($F_{3,33} = 8.917$, $p < 0.001$) but no main effect of run ($F_{2,22} = 2.858$, $p = 0.079$) and no interaction ($F_{6,66} = 1.982$, $p = 0.081$). These results suggest that participants were able to identify three out of four of the test conditions correctly from the beginning of the experiment, and that their performance did not appear to significantly improve as the experiment progressed.

Participants' response times revealed no differences between the consistent and violation conditions ($t_{11} = 1.311$, $p = 0.217$, Figure 4.4E). However, a RM-ANOVA with the factor 'sequence condition' (with levels 'familiar', 'novel', 'begins with A' and 'does not begin with A') revealed a significant main effect of sequence condition ($F_{3,33} = 3.425$, $p = 0.023$, Figure 4.4F). Bonferroni post-hoc tests do not show any differences between conditions, ($p > 0.1$ in all cases), however, Figure 4.4F suggests that this effect is driven by faster responses to violation sequences that 'do not begin with A' compared to all other conditions. Therefore, it appears that reaction time data only highlights that participants respond more quickly to those sequences with very early violations, while in other conditions responses were consistently slower.

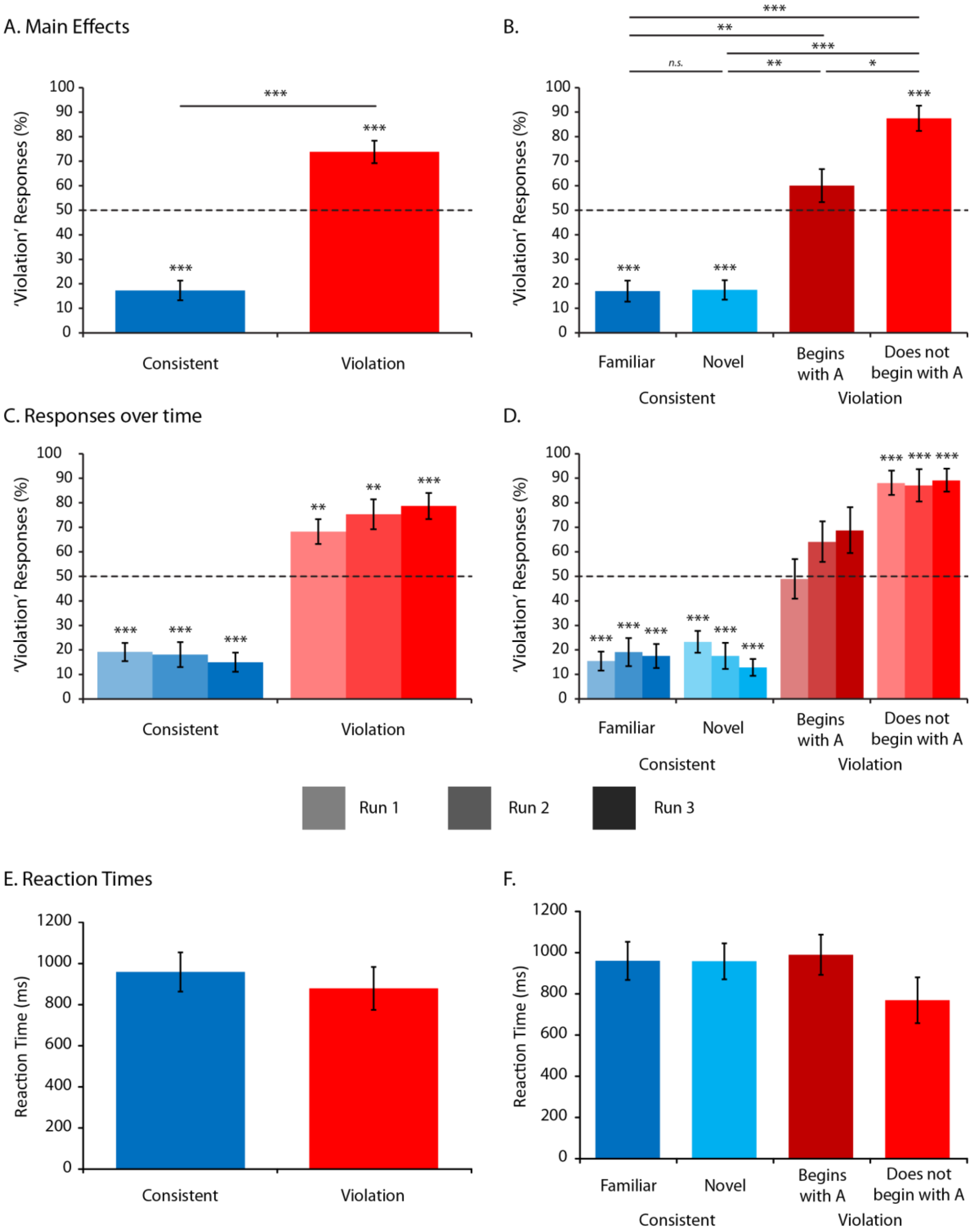


Figure 4.4 AGL results in human participants. (A-B) Mean proportion of trials on which participants responded 'violation' to test sequences (%, ±SEM). (C-D) Mean 'violation' responses split by testing run. (E-F) Mean response times separated by conditions. Significance levels are shown for conditions that are significantly different to chance (one sample *t*-tests), main effects (paired sample *t*-tests; left panels) and Bonferroni post-hoc tests (right panels); * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

4.5.3. Human conclusions

These results demonstrate that human participants are easily able to distinguish between consistent and violation sequences. Furthermore, most of the sequences were correctly classified at better than chance levels. However, participants did not classify violation sequences that begin with the A element significantly better than chance, suggesting that these sequences were harder to identify as violations. Furthermore, better performance on violation sequences that 'do not begin with A' compared to those that 'begin with A' show that violations in the initial sequence position were used as a significant cue in the identification of violation sequences. However, significant differences were observed between the violation sequences that 'begin with A' and both 'familiar' and 'novel' consistent sequences, demonstrating that participants did recognise differences between the consistent and violation conditions, even when they began with the same initial element. Conversely, no significant difference was observed between 'familiar' and 'novel' consistent sequences, suggesting that the familiarity of the test sequences did not influence participants' responses.

4.6. Discussion

The experiments in Chapter 3 demonstrated that Rhesus macaques (a species of Old World monkey) were able to implicitly learn a non-deterministic, forward branching artificial grammar. They produced more frequent, stronger and longer duration responses to sequences that violated the AG structure than those that were consistent with it. Furthermore, differences were observed between 'novel' consistent sequences and violations sequences that 'begin with A', but not between 'familiar' and 'novel' consistent sequences, or violation sequences that 'begin with A' and those that 'do not begin with A'. These results demonstrate that the artificial grammar learning observed in macaques cannot be attributed to simple strategies such as responding to previously unheard, novel sequences or only to violations in the initial parts of the test sequences.

By contrast, common marmosets (New World monkeys) showed no differences in the frequency or strength of responses to consistent and violation test sequences. However, their responses did persist longer (as evidenced by increased looking durations) following violation sequences. This result demonstrates that some level of AGL occurred in marmosets; the animals noticed the difference between consistent and violation sequences. However, unlike the macaques, no differences were observed between the 'novel' consistent sequences and the violation sequences that 'begin with A', which cannot be solved by simple strategies. Furthermore, while the data were insufficiently statistically robust to observe many differences between individual conditions, the pattern of marmoset responses appears to be different to that observed in macaques. The Rhesus macaques' responses showed a stepwise pattern of results, with no differences between the consistent ('familiar' and 'novel') or the violation conditions ('begins with A' and 'does not begin with A'), but a large difference between the consistent and violation conditions. By contrast, the common marmoset results appear to show a linear increase in response durations across the conditions from 'familiar' to 'does not begin with A'. Although the differences between individual conditions were not significant, this pattern of responses suggests that rather than the deeper learning evident in macaques, the marmosets' responses may well be based not only on the 'grammaticality' of a sequence, but also on the familiarity or novelty, or particularly on early violations in the test sequences.

Due to the small size and rapid movements of marmosets relative to macaques, the inter-rater reliability in this experiment was lower than that in the macaque video-coding experiment. However, more responses were recorded, and therefore included in later analyses, in the marmosets than the macaques; therefore differences between the species cannot be attributed to insufficient statistical power. It is possible that the marmosets' responses were more difficult to measure, thus this dataset contains more noise than the macaque data. However, even if additional data were to be collected in an effort to increase statistical power, while the variance of the data might decrease it is possible that the pattern of results would remain relatively consistent. Specifically, while reduced variance or increased statistical power in these analyses might reveal differences between the 'novel' consistent sequences and the violation sequences that 'begin with A', the results suggest that similarly large differences might be observed between 'familiar' and 'novel' sequences or violation sequences that 'begin with A' and 'do not begin with A'. Therefore, it is unlikely that the differences in results observed between species can be attributed to insufficient data in one species; these results suggest that marmosets appear use simpler cues than macaques. Furthermore, the macaque video-coding and eye-tracking experiments produced highly comparable results, therefore it is unlikely that the pattern of results would differ dramatically if an effective eye-tracking method were to be developed in common marmosets.

These conclusions share some similarities with previous nonhuman primate AGL studies. Saffran and colleagues, showed that cotton-top tamarins (a species of New World monkey) were able to discriminate between consistent and violation AG sequences, but only when the consistent test sequences were the same as those used in the habituation phase(Saffran *et al.*, 2008). Based on these results it is impossible to determine to what extent the tamarins' responses were driven by the novelty of the violation test sequences, rather than their knowledge of the structure of the AG. Furthermore, Fitch and Hauser (2004) showed that tamarins were able to learn the relatively simple $(AB)^n$ structure but not the more complex A^nB^n artificial grammar. In combination, these results might suggest that New World monkeys, including tamarins and marmosets, possess some AGL capabilities but these might be limited to simpler AG structures or relatively simple learning strategies.

Human participants, tested using an explicit version of this AGL paradigm, quickly learned to discriminate between consistent and violation sequences. Participants' response accuracies were also above chance for most test conditions. However, violation sequences that 'begin with A' were not correctly identified above chance levels, suggesting that these sequences were more difficult to accurately classify. It is important to note that while the humans' responses could be compared to chance levels, this was not possible in the nonhuman primates. With access only to the frequency, duration or strength of monkeys' responses, it is impossible to determine whether, for example, differences between conditions were caused by lower responsiveness to consistent sequences or increased responses to violation sequences, or how these would relate to chance performance in humans. While human participants failed to classify violation sequences that 'begin with A' above chance levels, significant differences between these and the 'novel' consistent sequences represent the same effect that is observed in macaques, but not in marmosets. Therefore, it is not possible to conclude that human participants were worse than macaques at recognising violation sequences that 'begin with A'. However, unlike the macaques, human participants responded differently to the violation sequences that 'begin with A' relative to those that 'do not begin with A'. Therefore, while we cannot conclude that humans were worse at recognising these violation sequences than macaques were, there is evidence that they at least use violations in the first element position when making their decisions. No difference was observed between the 'familiar' and 'novel' sequences. Therefore it appears that the participants did not assess the test sequences based on familiarity or novelty.

Like previous comparative studies in humans and nonhuman primates (Fitch & Hauser, 2004; Saffran *et al.*, 2008), it was not possible to test all the species using the same methods. Therefore, it may be difficult to establish whether any differences in results between humans and nonhuman primates could be attributed to the methods used rather than genuine cross-species differences. For example, humans, but not Rhesus macaques, showed different responses to the violation sequences based on the initial element. It is possible that this effect might be a result of human participants, after receiving explicit instructions, actively looking for rules or patterns governing the sequences. It is likely that participants may have recognised that all the habituation sequences began with the A element and that some of the test sequences did not. Some participants might then have failed to recognise more complex features of the AG structure, or assumed that they had worked out the rule and 'solved' the experiment. By contrast, the nonhuman primates could not be explicitly instructed, therefore likely based their responses, unconsciously, on some combination of the statistical regularities of the AG or on a broader combination of 'rules'. These experiments are insufficient to conclude that none of the differences observed between species might be attributable, at some level, to methodological differences. However, these results in humans, like many previous studies (e.g. Reber, 1967; Saffran, 2002; Folia *et al.*, 2010), demonstrated that human participants were quickly and accurately able to learn the AG structure. The results from Chapter 3 demonstrate that macaques were also able to learn this structure. It therefore seems reasonable to conclude that both testing methods present the different species sufficient opportunities to learn the AG. Therefore, it may be reasonable to compare the results gathered in humans with those from Rhesus macaques and common marmosets.

In combination, these results suggest an interesting pattern of AGL capabilities across primate species. Humans appear to learn the AG structure quickly, and easily discriminate between consistent and violation sequences. Macaques, following a longer habituation period, implicitly learn the structure to a high level, and their responses cannot be explained by simple strategies. Marmosets, under identical conditions to the macaques, show AGL effects only in the duration of their responses, and even these may be best explained by relatively simple strategies. These results may suggest that Rhesus macaques (Old World monkeys, which are more closely related to humans), may have better AGL capabilities than common marmosets, which are more distantly related New World monkeys. While data from two species is insufficient to draw broad conclusions about the abilities of New and Old World monkeys in general, they are at least suggestive of a potential evolutionary gradient of AGL abilities, with more closely related species possessing more complex sequence-structure learning abilities. Further research with more species would help to support or refute this hypothesis.

Human AGL abilities may be subserved by the same cognitive mechanisms and neuronal substrates that evolved for language processing (Friederici, 2004; Petersson *et al.*, 2004; Friederici, 2011; Petersson *et al.*, 2012). However, these evolutionary specialisations appear to have occurred much more recently than humans last shared common ancestor with either macaques or marmosets (e.g. Pinker & Bloom, 1990; Jackendoff, 2002). It is therefore difficult to determine what abilities or capacities such AGL experiments might be tapping into in a non-linguistic species like nonhuman primates. The comparative testing of different species may be able to shed light on this question (Hauser *et al.*, 2002; Petkov & Wilson, 2012). If a number of species were tested comparably on the same AG then it would be possible to analyse how AGL capabilities in these species might correspond to other, potentially related, differences between the species. For example, it is possible (although by no means necessary, Hauser *et al.*, 2002) that the same cognitive mechanisms that support AGL in nonhuman primates might also support their own vocal communication abilities. If this were the case, we might expect to observe more complex AGL abilities in those species with more complex communication systems, wider vocabularies of vocalisations, or those that vocalise more frequently. With only two species of nonhuman animal available for study, such analyses are difficult to perform. Both Rhesus macaques and common marmosets have relatively simple communication systems, largely relying on a relatively small number of vocalisations, produced either in isolation or series, but not in meaningful combination (Pola & Snowdon, 1975; Hauser & Marler, 1993b; 1993a). Marmosets however are an arboreal species, and as such produce frequent contact calls to maintain group cohesion in the forest canopy were vision is limited (DeVore, 1963). Rhesus macaques occupy a wide range of relatively open, savannah habitats where vocal communication is less important in maintaining group contact (e.g. Fooden, 2000). As more complex AGL capacities were observed in macaques, which possess either equal or possibly more limited vocal communication abilities, relative to marmosets, it seems unlikely that these AGL capabilities are subserved by the same abilities as the animals' vocal communication systems.

A wide range of systems or abilities might support the AGL abilities observed in primates. It is possible that the associations and relationships between nonsense words might recruit a truly generic system. This may, for example, naturally be involved in processing the social relationships between group members. Rhesus macaques live in relatively large troops with complex social structures and dominance hierarchies, while marmosets live in small groups or pairs (Hubrecht, 1984; Fooden, 2000). Therefore it is possible that these AGL abilities might correspond to the requirement to process and understand larger numbers of associations or relationships between individuals in more complex social groups. Alternatively, AGL abilities might relate to more general learning abilities. While data directly comparing macaque and marmoset abilities are uncommon, studies have shown that on a discrimination learning task macaques learn faster, and ultimately reach a higher level of performance, than marmosets do (Miles & Meyer, 1956; Miles, 1957). Therefore, it is possible that more complex AGL capabilities may be a correlate of these more efficient learning abilities in Rhesus macaques. These possibilities represent only two of a huge number of potential capacities that might be responsible for differences in AGL abilities across primate species. Data from a wider range of species, on both AGL capabilities and other, potentially related abilities, could help identify what cognitive mechanisms might support sequence-structure learning abilities in nonhuman primates. It may be particularly valuable to study, for example, species such as putty nosed monkeys, which have been shown combine their vocalisations in meaningful ways (Arnold & Zuberbuhler, 2006; 2008), to better clarify the relationship between AGL abilities and the complexity of an animals' vocal communication system. Data from two species are insufficient to draw firm conclusions, however, these results minimally allow us to suggest that macaques, which are both more closely related to humans and possess larger brains than their New World cousins, common marmosets, appear to have more complex AGL capabilities.

4.7. Conclusion

The results of these experiments demonstrate that the AGL capabilities previously observed in Rhesus macaques are also present in humans, and to some extent, in at least one other nonhuman primate species, and are therefore likely to be present, in some form or other, in a range of primate species. Furthermore, although the mechanisms supporting these abilities are not clearly understood, these results seem to suggest that more closely related species might perform better at these tasks than more distantly related ones. Whether or not this result generalises more broadly to other Old and New World monkeys, it may be possible to conclude that while common marmosets do show some limited AGL abilities, Rhesus macaques might represent the stronger candidate for an animal model system in which to study language related processes in the brain.

These results are able to inform us about the relative sequence-structure learning capabilities of two nonhuman primate species, and may, by inference, begin to provide some information about the abilities of nonhuman primates more generally. However, these, and any behavioural experiments, are likely to be insufficient to address whether the mechanisms involved in AGL in nonhuman primates represent homologues of some aspects of the language processing system in humans, which would suggest a generic basis for this system, or some other, more distantly related process. Therefore, in the next chapter, I will present results from comparative fMRI experiments in Rhesus macaques and human participants, which aimed to assess the similarities and differences between the brain areas involved in AGL in the two species, and how this might relate to the language network in the human brain.

Chapter 5: Artificial Grammar Learning in the Primate Brain

5.1. Abstract

Human neuroimaging studies have shown that natural language processing engages left hemisphere perisylvian brain regions. Artificial Grammars (AGs) can be designed to emulate aspects of the syntactic structure of language, and have been shown to produce activity in comparable brain areas. Nonhuman primates have been shown to learn some AGs. Data regarding the brain regions that support these processes in nonhuman animals have the potential to provide important insights into the evolution of language related abilities. However, such data was not previously available. Functional magnetic-resonance imaging (fMRI) experiments were performed in Rhesus macaques and human participants, which showed that several perisylvian regions are comparably engaged in AG learning in both species. Corresponding sensitivity to violations of the AG were observed in the inferior frontal gyrus, ventral to presumed homologues of Broca's area (BA44/45), as well as temporal and temporo-parietal regions. BA44/45 was statistically involved in the macaques, but not in humans, suggesting interesting implications about the evolution of Broca's territory in humans. Furthermore, rather than being left lateralised, the effects were bilaterally distributed in both species. The observed correspondences show that humans process this AG using an evolutionarily conserved set of perisylvian brain regions whose function involves evaluating learned sequencing relationships. Such processes reflect domain-general rather than language-specific functions, the latter of which in humans might depend more on left lateralised processes and on Broca's territory. These corresponding brain areas can now be studied in animal models at a neuronal level.

5.2. Introduction

Language related processes recruit frontal, temporal and parietal brain regions that surround and reside in the lateral sulcus or Sylvian fissure ('perisylvian' regions, Broca, 1861a; Damasio & Geschwind, 1984; Binder *et al.*, 1997; Catani *et al.*, 2005; Hickok & Poeppel, 2007). Neuroimaging studies have been used in combination with Artificial Grammar (AG) learning paradigms to evaluate the brain regions involved in syntax-related processes, independently of semantic processes (Petersson *et al.*, 2004; Friederici *et al.*, 2006a; Friederici, 2011; Petersson *et al.*, 2012). These studies have highlighted the involvement of the left inferior frontal gyrus, including Broca's area and more ventral regions, as well as parts of the temporal lobes and temporo-parietal regions. Previous experiments, including those reported here, have demonstrated that nonhuman animals are able to learn AGs of various levels of complexity (Chapters 3 and 4 and, e.g. Fitch & Hauser, 2004; Gentner *et al.*, 2006; Murphy *et al.*, 2008; van Heijningen *et al.*, 2009). However, which brain regions support such processes, and how these compare to human brain regions involved in similar processes, was previously unknown.

Identifying the nonhuman primate brain regions involved in AGL is an important goal. A number of scenarios regarding which areas are recruited are possible. Firstly, it is possible that behavioural similarities observed between humans and nonhuman animals might be supported by different cognitive mechanisms and brain areas, because humans have access to a specific language processing system which is not present in nonhuman animals. Alternatively, AGL paradigms might activate comparable, homologous regions in both humans and nonhuman primates. This would suggest that the brain areas that support language in humans might share a common evolutionary heritage with regions observed in extant nonhuman primates, suggesting a more generic evolutionary basis for aspects of language processing. Furthermore, the identification of functionally homologous brain regions in macaques would represent a critical step forward in the development of an animal model system in which language related abilities could be tested at a neuronal level. Conversely, evidence that AGs are processed by discrete sets of brain regions in humans and monkeys would suggest that the value of comparative AGL paradigms as a method for investigating language evolution might be more limited. It is also possible that both correspondences and differences might exist between the species. For example, some functional homologies might be observed, but other features, such as the lateralisation of these areas observed in the human language network, might not be present in non-linguistic species such as nonhuman primates.

If comparable perisylvian brain areas were activated by the same AG in both species, the specific patterns of activity observed could provide valuable insights into the evolution of specific, brain regions involved in language processing in humans (Friederici, 2002; 2004; Petkov & Wilson, 2012). Human AGL neuroimaging experiments suggest that simple syntax-building functions, such as those involved in evaluating adjacent relationships in finite-state Artificial Grammars (Chomsky, 1957), engage ventral frontal cortex (vFC) regions, including frontal-opercular areas (Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009; Friederici, 2011). However, Broca's territory (left Brodmann areas 44 and 45) does not appear to be strongly involved in these initial processes, unless more complex, non-adjacent AG relationships or syntactic and semantic features of natural language are involved (Friederici *et al.*, 2006a; Tyler *et al.*, 2011; Petersson *et al.*, 2012). Primate AGL studies have shown that nonhuman primates (cotton-top tamarins) were able to learn a simple AG, i.e. $(AB)^n$, which produces activation in vFC regions in humans. However, they were unable to learn a more complex A^nB^n structure, reported to recruit Broca's area in humans (Fitch & Hauser, 2004; Friederici *et al.*, 2006a). Furthermore, this simpler syntactic processing relies upon a ventral pathway between the anterior temporal lobe and vFC, while more complex AGs recruited a dorsal pathway between BA44/45 and posterior temporal regions (Friederici *et al.*, 2006a; Bahlmann *et al.*, 2009). DTI experiments have identified evidence of the ventral pathway in Rhesus macaques and chimpanzees, however evidence for the existence of a dorsal pathways is much weaker in these species (Rilling *et al.*, 2008). Therefore, whether AGL paradigms might activate structural homologues of Broca's area (BA44/45) in nonhuman primates, or any structures implicated in the dorsal pathway in humans, remains an important question. It is possible that the inability of cotton-top tamarins to learn the more complex A^nB^n AG (Fitch & Hauser, 2004) may be due to Broca's area representing a unique specialisation for language, which is not present in nonhuman species. Alternatively, since the processing of complex, finite-state AGs has been shown to activate BA44/45 in humans (Petersson *et al.*, 2004; Petersson *et al.*, 2012), the present AG structure might represent a sufficient level of complexity to cause activation in this area in nonhuman primates, but not necessarily in humans.

This functional magnetic-resonance imaging (fMRI) study evaluated which brain regions in humans and Rhesus macaques are sensitive to a non-deterministic, forward branching AG. Insights on anatomically and functionally corresponding regions across the species, which are sensitive to the AG, could provide evidence to clarify conserved and specialised human function and support future neuronal-level studies of the conserved regions in animal models.

5.3. Rhesus macaque fMRI experiment

Behavioural experiments in both Rhesus macaques and common marmosets showed that these species were sensitive to AG violations (see Chapters 3 and 4). However, these effects were much more robust in macaques than marmosets, and could not be explained by simple strategies. Furthermore, macaques are an Old World monkey, more closely related to humans than marmosets, and therefore might represent the more appropriate animal model system. Finally, macaques are larger animals that are easier to test in neuroscientific experiments, including both fMRI experiments and potential future neurophysiological recordings. Therefore, in this experiment, three Rhesus macaques participated in an implicit AGL task during fMRI scanning.

5.3.1. Methods

Stimuli

The stimuli used were identical to those used in the previous experiments (see Section 3.4.1 and Figure 3.1).

Ethics Statement

All animal work and procedures were approved by the U.K. Home Office and abide by the Animal Scientific Procedures Act (1986) on the care and use of animals in research.

Participants

Two of the Rhesus macaques (*Macaca mulatta*) who participated in both the eye-tracking experiment and the video-coding experiment, and one who only participated in the video-coding experiment (see Chapter 3) were tested using a sparse scanning fMRI paradigm. Due to practical and ethical considerations it is typically not possible to scan large numbers of macaques in the same way human fMRI experiments are conducted. Therefore, analyses were performed individually in each animal and the results are discussed with specific focus on areas of activation that were consistent across the majority of animals (2 out of 3). All the animals were male, weighing 6-12kg, and lived in a colony of pair housed macaques. They were trained to complete trials of a sparse-imaging sequence with visual fixation and minimal body movement during stimulation (Petkov *et al.*, 2006; Petkov *et al.*, 2008a).

FMRI experiments require the heads of nonhuman primates to be immobilised to acquire high quality data and to prevent movement artefacts. This is achieved by means of a surgically implanted MR-compatible plastic head-post, secured to the skull with ceramic screws and dental acrylic. This procedure is common in nonhuman primate neuroscientific research; however, like any surgical procedure, problems sometimes occur. Throughout the course of this research, two of the animals scanned began to reject, and ultimately lost their head-posts (Monkeys 1 and 3). This made fMRI scanning impossible in these animals until the animals have sufficiently recovered to be re-implanted. Due to these medical issues, it was impossible to collect identical amounts of data in each animal. The animals participated in 1, 8 and 5 scanning sessions respectively. Scanning sessions consisted of an average of 5 testing runs including 30 testing trials each. Despite the differences in the amount of data recorded in each animal, analyses were performed identically in each case (see Rhesus macaque fMRI data analysis, below). All data were cluster corrected ($p < 0.05$ level), and all three animals showed significant cluster corrected activity responses, including several clusters for Monkey 1 who had the fewest trials (Table 5.1). Furthermore, the pattern of effects observed between the animals did not appear to be closely related to the amount of data available for analysis.

Auditory cortex (AC) maps were individually defined for each animal based on methods described in Petkov *et al.,* (Petkov *et al.*, 2006). For each animal, the average BOLD activation to stimulus sequences (regardless of whether they were consistent or violation sequences) relative to silent trials was calculated, across both ACs. These data were then compared to the volume of scanning data obtained. Monkeys 1 and 2, in which 1 and 8 scanning sessions were recorded respectively, showed comparable, high activation in AC (mean z-score of 0.83 and 1.01, averaged across the entire left and right auditory cortices). Monkey 3, in which an intermediate number of sessions (five) were recorded showed much lower AC activation (mean z-score of 0.18). No correlation was apparent between number of sessions and AC activity ($r = 0.11$). Furthermore, a similar analysis was performed to asses any possible relationship between number of scanning sessions and activation to violation relative to consistent sequences in the ventral frontal cortex (vFC, see Rhesus macaque region of interest analyses, below). Again higher activation was observed in Monkeys 1 and 2, while Monkey 3 showed lower activation ($r = -0.30$). These analyses suggest that the number of scanning sessions recorded do not correspond strongly with either levels of activation in general (as measured by the sound-silence contrast in the AC) or with activation to the violation-consistent sequences in key brain areas. While collecting more balanced datasets would have been desirable, medical and logistical issues made this impossible. Nevertheless, these analyses suggest that the results of this experiment cannot easily be attributed to these differences.

Habituation phase

Each scanning session began with a 30 minute habituation phase during which the habituation sequences were presented binaurally in a randomised order (rate of 9 sequences/min; inter-sequence interval = 4 sec; Figure 3.1).

Test phase

In the scanner, the monkey was encouraged to fixate on centrally located fixation spot for a juice reward (delivered after the scanning trial completed). Following each fixation period the animal was presented with a randomly selected consistent or violation test sequence (40% each of trials) or a silent trial (20% of trials). Stimuli were presented binaurally through MR-compatible headphones (Nordic Neurolabs) using Cortex software (Salk Institute) at ~75 dB SPL (calibrated with an XL2 sound level meter, NTI Audio). Each scanning run consisted of 30 trials (24 stimulus trials, 3 repeats of each sequence, Figure 3.1), following which the monkey was re-exposed to the habituation sequences for 5 minutes. Up to six scanning runs were performed each day, based on the animal's willingness to engage the fixation spot for a juice reward.

Rhesus macaque magnetic resonance imaging

Measurements of the fMRI blood oxygen level dependant (BOLD) signal were made with a nonhuman primate dedicated, vertical bore 4.7 tesla scanner (Bruker BioSpin) at Newcastle University. The monkeys sat in a primate chair in the scanner. Signals were acquired using a birdcage RF coil. Functional data were acquired using a single-shot gradient-recalled echo planar imaging (GE-EPI) sequence. The use of a sparse scanning paradigm allowed the stimulus sequences to be presented in relative silence (Petkov *et al.*, 2009). Each trial involved the acquisition of a baseline volume (volume acquisition time 2s), followed 9 seconds later by a stimulus volume (echo spacing = 9s; flip angle: 90^{o}; TE = 22ms; 16 slices, 2mm; in-plane field of view: 12.8 x 9.6cm^2, on a grid of 128 x 96 voxels, with voxel resolution of 1 x 1 x 2mm^3; inter-trial-interval = 9s; Figure 5.1). On stimulus trials (80% of trials) a randomly selected stimulus sequence was presented aligned so that the stimulus ended 2 seconds before the stimulus volume began (stimulus onset occurred on average 3.8 seconds after the start of the initial volume, stimulus offset occurred at 7 seconds, Figure 5.1). These timings were developed to measure the peak BOLD response produced by the stimulus relative to the initial baseline volume. The BOLD response to auditory stimuli peaks approximately 4 seconds following stimulus presentation (Baumann *et al.*, 2010), therefore this scanning paradigm was designed to maximise response to the stimuli while minimising activation caused by the preceding volume (Figure 5.1). Anatomical images were acquired in register with functional scans in each scanning session using a 3D T1-weighted MDEFT sequence with parameters TE: 6ms; echo spacing: 750ms; inversion delay: 700ms; 22 slices; in-plane field of view: 12.8 x 9.6cm^2, on a grid of 256 x 192 voxels, with voxel resolution of 0.5 x 0.5 x 2mm^3, number of segments: 8.

Rhesus macaque fMRI data analysis

For each animal, first-level general linear model (GLM) analyses with fixed effects (FEAT, FSL) were performed, contrasting BOLD responses to violation-consistent sequences. Functional data for each animal was registered to a template functional image rather than a high resolution anatomical or standard image. This approach allowed the functional data for each animal to be compared in a common reference space, ensuring that areas of activation could be accurately compared between scanning runs. Only significant clusters ($p < 0.05$, cluster corrected) that occurred in corresponding voxels or anatomical regions (Saleem & Logothetis, 2007) in at least the majority of the three animals were discussed (Table 5.1 and Figure 5.2). The data were then registered to a standard reference space for comparison (Saleem & Logothetis, 2007), and projected to surface-rendered standard template brain using FreeSurfer for clearer display purposes (McLaren *et al.*, 2009). Consistent-violation contrasts were also performed. However, these analyses produced no significantly active clusters in one of the animals, and no correspondences between the clusters observed in the other two. Furthermore, the previous literature has generated no strong *a priori* hypotheses about the regions that might be expected in this contrast. Therefore, these contrasts were not considered further.

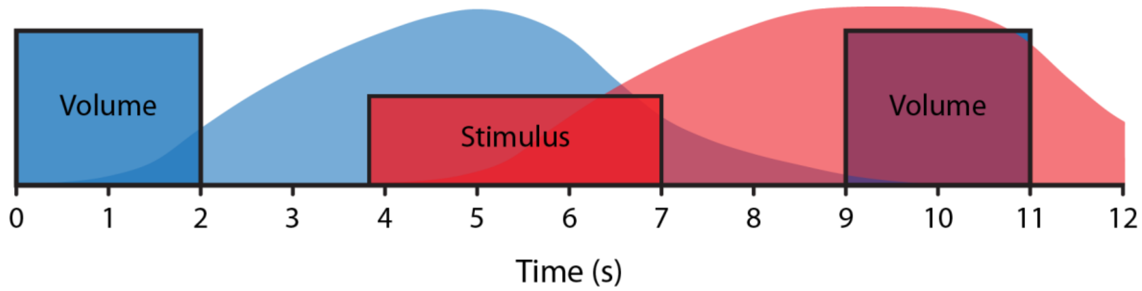


Figure 5.1 Scanning paradigm and predicted BOLD responses. Initial stimulus volume begins at trial time = 0s. Stimuli were individually aligned to end 2 seconds before the second, stimulus volume, which began at 9s. Schematics of predicted BOLD responses to the first volume (blue) and the stimulus sequence (red) are also shown (based on the time-course of BOLD signal to auditory stimuli Baumann *et al.*, 2010). Timings were designed to maximise activation in response to the stimulus relative to the initial volume during the second, stimulus volume. Vertical heights are arbitrary and only meant for illustrative purposes.

Rhesus macaque region of interest analyses

A BA44/45 Region of Interest (ROI), representing a potential structural homologue of Broca's area in humans was anatomically defined, consisting of Brodmann areas 44 and 45 (Saleem & Logothetis, 2007). Following imaging studies in human participants and previous hypotheses (see Chapter 1, Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009; Petkov & Wilson, 2012) a ventral Frontal Cortex (vFC) ROI was anatomically defined, including precentral opercular areas (PrCO), ventral BA6 and the dysgranular insula (Figure 5.3G). Somatosensory areas 1, 2 and 3, the gustatory cortex, and BA12 were excluded from this ROI based on *a priori* hypotheses, despite their anatomical proximity to these vFC areas (see Introduction, Friederici *et al.*, 2006a; Petkov & Wilson, 2012 Figure 5.3G). These ROIs were mapped in a standard reference space individually in both hemispheres. The average BOLD signal in response to violation–consistent contrasts was calculated for each voxel in each ROI. These values were normalised based on the maximum activation for each animal, to ensure that analyses revealed relative differences between violation and consistent sequences, rather than overall stronger activation in any individual animal. RM-ANOVAs and Bonferroni corrected one-sample *t*-tests assessed whether the ROI was significantly activated by violation sequences relative to consistent sequences. RM-ANOVAs and two-sample *t*-tests were used to investigate lateralisation by comparing activation across hemispheres. RM-ANOVAs were used to compare the macaque data to those results obtained in humans.

5.3.2. Results

After habituation to exemplary AG sequences (Figure 3.1), the three macaques were scanned with functional magnetic-resonance imaging (fMRI) as they listened to consistent or violation testing sequences (Figure 3.1, Methods). Brain regions sensitive to violations of the AG for each animal (contrast: 'violation' vs. 'consistent') were mapped onto a surface-rendered standard template brain (Figure 5.2A-C, McLaren *et al.*, 2009). An overlap map showing the average BOLD response of the voxels were activated in at least 2 out of 3 of the monkeys is also shown (Figure 5.2D; also Table 5.1).

These GLM analyses showed activation in a number of perisylvian brain regions, including some broad consistencies across different animals. In particular, ventral portions of the frontal cortex were involved in all three animals. In the first monkey the activation includes a large portion of the IFG (including BA6), extending from the dysgranular insula dorsally through the ventral frontal cortex up to BA44 and BA45. The second animal shows a comparable pattern, although BA44 and BA45 appear to be relatively less involved, while this cluster of activation extends ventrally into the anterior temporal lobe (including area TS2). The third animal shows strong activation in the dysgranular insula and anterior temporal lobe, although some more dorsally located IFG regions did not reach cluster corrected significance levels. These GLM results suggest that ventral portions of the IFG and temporo-frontal regions including the dysgranular insula, are critically involved in AGL in the primate brain. Posterior temporal and temporo-parietal regions (Area 7) were also observed in a majority of the monkeys. This area represents a similar anatomical region to the angular gyrus (BA39) in humans, which is also known to be involved in language processing (see Discussion, Friederici *et al.*, 2003; Friederici *et al.*, 2010). Finally, corresponding activation was seen in the caudate nucleus in two of the animals, but was not obvious in the majority consensus map because the significantly active voxels were not in identical locations, but were in comparable anatomical areas.

These GLM analyses appear to suggest that greater activation in response to the violation sequences relative to the consistent sequences occurred in the right hemisphere, at least in some of the animals. However, these analyses used a relatively conservative, cluster-corrected significance threshold (see Methods). Activation in each voxel was assessed relative to an initial significance threshold, and only clusters of significant voxels above a certain size appear in the final analyses (Fig. 5.2D, Table 5.1). Therefore, it is possible that these analyses might under-represent left hemisphere activation, which may be present, but below the significance level observed in the right hemisphere. To conclude that the effects seen here are statistically right lateralised, beyond identifying supra-threshold activation in one hemisphere and sub-threshold activation in the other, it is necessary to demonstrate that there was significantly greater activation in the right than the left hemisphere. To directly compare the BOLD signal across hemispheres and between specific regions within the IFG, Region of Interest (ROI) analyses were performed. ROIs for BA44/45 and more ventral Frontal Cortex (vFC) regions were anatomically defined, and activation in these regions was quantitatively assessed.

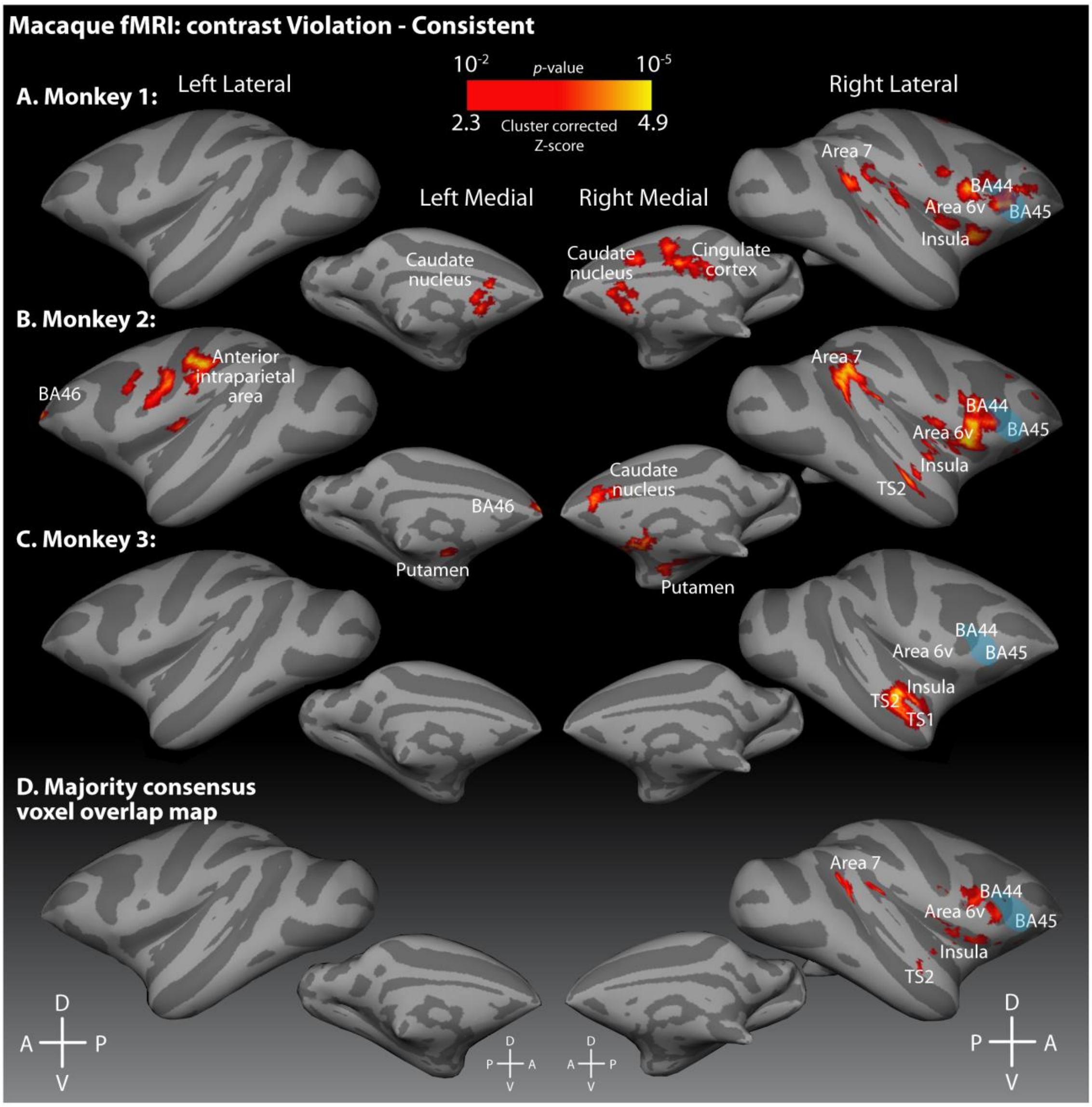


Figure 5.2. Macaque brain regions sensitive to AG violations. Statistical parametric maps of sensitivity to AG violations (contrast: violation vs. consistent) displayed in each of the three macaques (A-C) and in a majority consensus voxel-overlap map (D), all $p < 0.05$ cluster corrected. Results are displayed on rendered medial and lateral surface representations transformed to a standard monkey brain (McLaren *et al.*, 2009) that is in register with an accepted atlas of the identified anatomical regions in macaques (Saleem & Logothetis, 2007). Abbreviations: D = Dorsal; V = Ventral; A = Anterior; P = Posterior.

Table 5.1 Anatomical locations of significant clusters in macaques. Coordinates (in stereotactic space) and maximum *z*-scores for significantly activated clusters, in each of three Rhesus macaques, including locations where activation co-occurs in at least two out of three monkeys.

Anatomical Location	**Stereotactic coordinates**			**Cluster corrected max. *z*-score**	**Hemisphere**
	x	*y*	*z*		
Monkey 1					
Ventral Frontal Cortex	25	26	20.5	3.55	Right
BA45, BA44, dysgranular insula, BA6, putamen					
Temporo-parietal regions	14.5	8	17.5	4.16	Right
BA7, Caudate nucleus (posterior), posterior auditory cortex (areas Pro, CM)					
Cingulate cortex	8	14.5	25	3.08	Right
Anterior/posterior cingulate cortex, caudate nucleus					
Caudate nucleus	1.5	24	18	3.08	Bilateral
Monkey 2					
Ventral frontal cortex & anterior temporal lobe	27	21.5	11.5	4.89	Right
BA6v, BA44, BA45, dysgranular insula, anterior auditory cortex (area TS2)					
Frontal pole (BA46)	-4	44.5	20	3.93	Left
Anterior intraparietal area	-19.5	12.5	29.5	3.88	Left
Anterior and ventral intraparietal areas					
Caudate nucleus & putamen	5	32.5	20.5	3.87	Right
Temporo-parietal regions	16	1.5	25	3.52	Right
BA7, posterior auditory cortex (area TPT)					
Nucleus accumbens	2.5	20.5	7	3.49	Bilateral
Auditory cortex and insula	22	16	10.5	3.10	Right
Auditory cortex (areas R, RM), dysgranular insula, putamen					
Monkey 3					
Temporal cortex & insula	26	19	7.5	3.60	Right
Anterior auditory cortex (areas TS1, TS2), dysgranular insula					
Majority consensus (voxels activated in ≥2 monkeys)					
Anterior auditory cortex (area TS2)	26.5	20.5	11.5		Right
Dysgranular insula	20.5	17	17		Right
BA6v	24.5	27	19.5		Right
BA7	12	0.5	23		Right
Caudate nucleus & putamen	14.5	24	17		Right

ROI Results

To evaluate the involvement of BA44/45 and adjacent ventral frontal cortex (vFC) regions in response to AG violations, the following planned ROI analyses were conducted. Separate ROIs were anatomically defined bilaterally for, Brodmann areas 44 and 45 (BA44/45, blue in Figure 5.2G); and adjacent ventral regions including ventral frontal-temporal opercular regions, (vFC, green in Figure 5.2G, see Methods).

For each monkey, a voxel-based Repeated-Measures (RM) ANOVA was used to evaluate effects in the ROIs, with the factors 'condition' (consistent, violation), 'hemisphere' (left, right) and 'ROI' (FOP, BA44/45). All three animals showed a main effect of 'condition' (M1, $F_{1,1143} = 62.7; p < 0.001$; M2: $F_{1,1918} = 41.5; p < 0.001$; M3: $F_{1,1825} = 5.219; p = 0.02$) demonstrating that across both ROIs, all three monkeys showed increased activation to violation sequences relative to consistent ones, suggesting that these regions are critically involved in processing this AG. One monkey showed a significant interaction between 'condition' and 'ROI' (M2: $F_{1,1918} = 17.191$; $p < 0.001$) demonstrating increased activation in the vFC relative to BA44/45, however in the other two animals, this failed to reach statistical significance (M1: $F_{1,1143} = 2.469; p = 0.116$; M3: $F_{1,1825} = 3.838; p = 0.051$). Finally, lateralisation was assessed by means of an interaction between 'condition' and 'ROI'. Monkeys 1 and 3 showed no lateralisation effects (M1: $F_{1,1143} = 0.860; p = 0.860$; M3: $F_{1,1825} = 0.235$; $p = 0.628$), suggesting that these ROIs are equivalently engaged in both hemispheres, however, in Monkey 2 significant right-lateralisation was observed ($F_{1,1918} = 19.521$; $p < 0.001$). To assess the effects in the vFC and BA44/45 ROIs independently, additional analyses in each monkey were performed.

All animals showed significant vFC activation in response to violation sequences relative to consistent sequences in the right hemisphere, and this difference was also present in the left hemisphere in two out of the three animals (Bonferroni corrected one-sample *t*-tests, M1, left: $t_{404} = 4.155, p < 0.001$; right: $t_{366} = 3.827, p < 0.001$; M2, left: $t_{702} = 12.052, p < 0.001$; right: $t_{658} = 2.401, p = 0.034$; M3, left: $t_{658} = 3.818, p < 0.001$; right: $t_{632} = 1.523, p = 0.256$; Figure 5.2A-C). No lateralisation effect was observed in two of the three animals (two-sample *t*-tests: M1: $t_{770} = 1.279, p = 0.201$; M3: $t_{1290} = 1.839, p = 0.066$; Figure 5.2A&C) but the effect was stronger in the right hemisphere of Monkey 2 (two-sample *t*-test: $t_{1360} = 5.771, p < 0.001$; Figure 5.2B).

In the BA44/45 ROI, significant activation was revealed in two out of the three monkeys (M1 left: $t_{192} = 8.000, p < 0.001$; right: $t_{181} = 2.438, p = 0.032$; M2: left: $t_{276} = 0.05$,

$p = 1.0$; right: $t_{282} = 2.372$, $p = 0.036$; Figure 5.2D&E). Monkey 3 showed no activation in BA44/45 (left: $t_{265} = 0.583$, $p = 1.0$; right: $t_{270} = 0.331$, $p = 1.0$, Figure 5.2F). None of the monkeys showed any lateralisation effect between the BA44/45 ROIs (M1: $t_{373} = 0.977$, $p = 0.329$; M2: $t_{558} = 1.556$, $p = 0.120$; M3: $t_{535} = 0.635$, $p = 0.514$; Figure 5.2D-F).

These results demonstrate that the ventral frontal cortex is highly sensitive to violations of the AG, showing significant activation relative to consistent sequences in all animals tested. Although the pattern is less consistent, Brodmann Areas 44 and 45, which are putative structural homologues of human Broca's area, also appear to be involved in at least two of the animals. Interestingly, despite the appearance of the GLM analyses, no lateralisation was observed in two out of three macaques, although one monkey did show increased activation in the right hemisphere, particularly in the ventral frontal cortex.

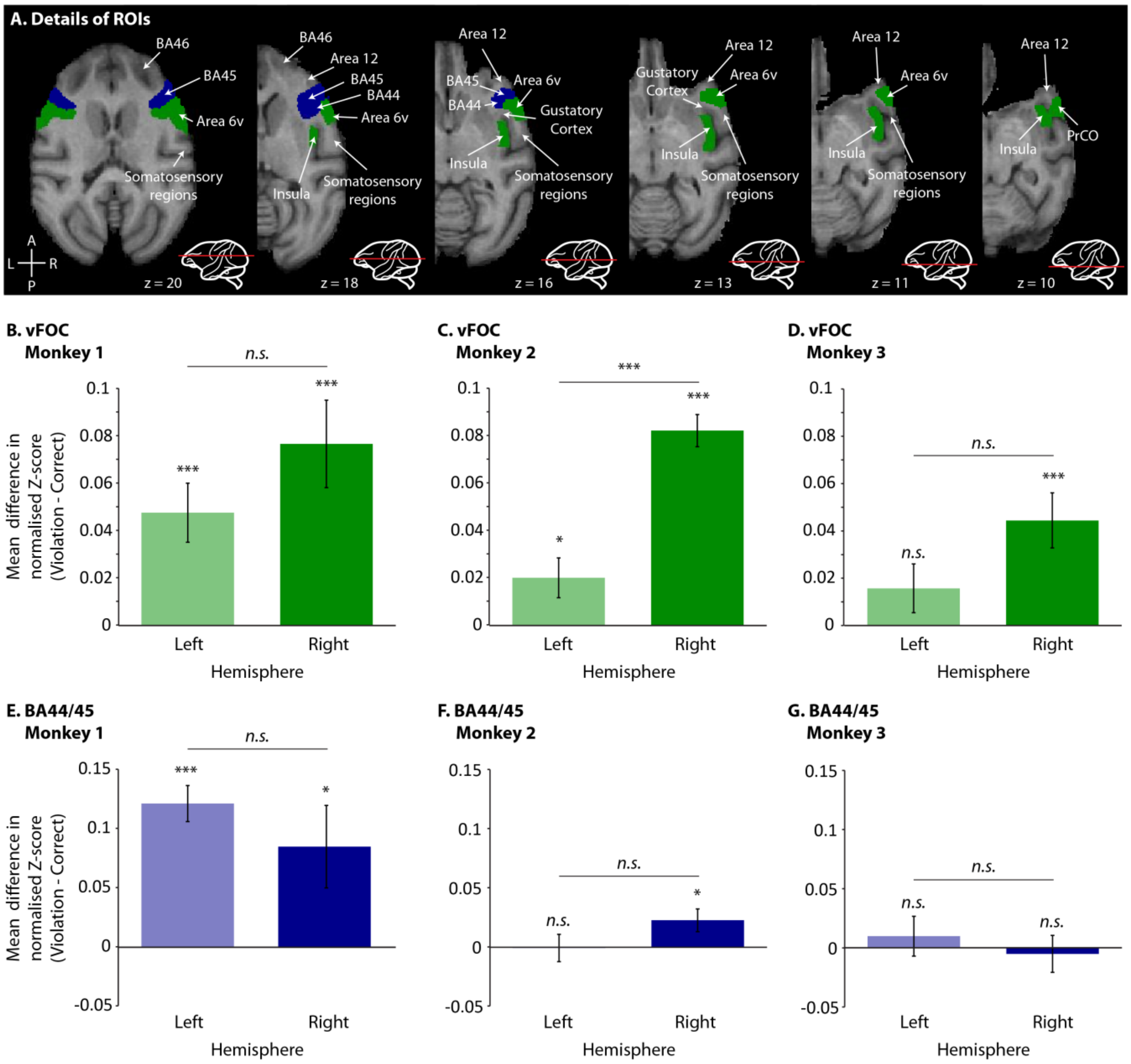


Figure 5.3. Macaque ROI results. (A) Anatomically defined bilateral ROIs used for analyses: blue comprises Brodmann areas 44 and 45 (BA44/45); green comprises adjacent ventral frontal (vFC) regions, including anatomical areas PrCO, dysgranular insula and ventral BA6. Somatosensory and gustatory regions and Area 12 were excluded from analysis (see Methods). (B-G) Normalised mean BOLD response differences by voxel (violation vs. consistent) for BA44/45 and vFC in the left and right hemisphere of the three macaques. Abbreviations: A = Anterior; P = Posterior; L = Left; R = Right; Bonferroni corrected *t*-tests: * = $p < 0.05$; *** = $p < 0.001$.

5.3.3. Rhesus Macaque conclusion

The results of this fMRI experiment demonstrate that a number of perisylvian areas in the macaque brain are sensitive to violations of this non-deterministic AG. Ventral frontal cortex and temporo-frontal regions including the dysgranular insula showed significantly stronger activation in response to violation than consistent sequences in all three animals. Additional regions including the anterior temporal lobe and temporo-parietal Brodmann area 7 were also strongly activated. These regions correspond with a number of structurally homologous regions involved in the processing of language and AGs in the human brain (e.g., Friederici, 2011). Furthermore, ROI analyses showed that ventral frontal cortex regions were strongly sensitive to violations of the AG in all animals, while BA44/45, which is a presumed structural homologue of the human Broca's area, was activated less consistently across animals. Interestingly, two of the macaques showed no significant lateralisation effect, although one animal did show significant right hemisphere lateralisation in the vFC. However, such results, in a minority of the animals, are insufficient to conclude that processing finite state artificial grammars, such as the one used here, primarily recruits the right hemisphere. These results have interesting implications regarding how areas of the human brain may have specialised for language, and which regions may be functionally conserved from a common ancestor. However, to draw firm conclusions regarding these results, and to assess the appropriateness of the Rhesus macaque as an animal model in which aspects of language might be studied, it is necessary to directly compare these results to those of human participants tested with an identical AG.

5.4. Human fMRI experiment

Neuroimaging experiments can provide valuable insights into the brain areas that support AGL in a nonhuman primate, and therefore potentially into the evolution of these brain areas and the abilities that they support. However, to fully understand the implications of these results, it is necessary to compare them directly to data from human participants, tested with the same AG structure. While a large number of human fMRI AGL experiments have been conducted (e.g. Petersson *et al.*, 2004; Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009; Petersson & Hagoort, 2012), no previous study has used this specific artificial grammar. While AGL paradigms using different AG structures show some consistent results between experiments, there is evidence that different brain areas are activated as a function of the complexity of an AG structure (Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009), including finite-state grammars similar to the one used here (Petersson *et al.*, 2004; Petersson & Hagoort, 2012). Therefore, to ensure that the results obtained in humans and macaques are as comparable as possible, it was necessary to test humans using the same AG structure.

While the Rhesus macaques could not be informed about the nature of the experiment, and were tested using entirely implicit measures, a different approach was required in the human experiment. The aim of the behavioural experiment in Chapter 4 was to confirm that human participants were able to learn the AG structure, and to inform us about the strategies they might use. The goal of this fMRI experiment was to investigate which brain areas might be involved in the detection of violations of the AG structure. Therefore participants were selected based on their performance on the prior behavioural experiment. This ensured that the participants (like the macaques who had all participated in at least one previous behavioural experiment, see Chapter 3), were able to learn the AG. In addition to being provided with instructions about the nature of the experiment (although not about the rules or structure of the AG), as in the behavioural experiment, participants were asked to respond after each stimulus presentation (see Chapter 4) in order to maintain attention throughout the fMRI experiment.

Finally, it was not possible to collect a sufficient amount of data in individual human participants to perform identical analyses to those used in the macaques. Therefore, 9 participants were each scanned once. Statistical analyses were conducted in as similar a manner as possible (see Methods), however insufficient data was available to compare between individual human participants.

5.4.1. Methods

Stimuli

The stimuli used were identical to those used in the previous experiments (see Section 3.4.1 and Figure 3.1).

Participants

Nine adult humans (age range 19 to 29 years, mean age 22; 3 male, 6 female) who had previously participated in the behavioural AGL experiment (see Chapter 4) participated in this fMRI experiment. Participants were recruited through the Newcastle University Institute of Neuroscience participation scheme and provided informed consent to participate. All participants were native English speakers, had normal hearing and normal or corrected vision. No participants had any language or comprehension disorders. The ethics of this experiment were approved by the local ethics committee at Newcastle University.

Habituation phase

Participants were exposed to the habituation sequences in a random order for 10 minutes, prior to the start of scanning (rate of 9 sequences/min; inter-sequence interval = 4 sec; Figure 3.1).

Test phase

Each scanning trial began with the presentation of a blue fixation spot. A randomly selected consistent or violation stimulus sequence (40% each of trials, Figure 3.1) or a silent trial (20% of trials) was then presented binaurally through MR compatible headphones using custom Matlab scripts and Psychophysics toolbox. Following the stimulus offset the fixation spot changed to yellow and, as in the behavioural experiment, the participants were asked to use a response box to respond whether the sequence followed the 'same' or 'different' pattern or rules to the habituation sequence. During fMRI scanning, participants performed at a high level, accurately classifying 87.2% of the testing sequences. Each scanning run consisted of 50 trials (40 stimulus trials, 5 repeats of each sequence, Figure 3.1), following which the participants were re-familiarised with the habituation sequences for 3 minutes. All participants took part in 4 scanning runs within the scanning session.

Human magnetic resonance imaging

Measurements of the fMRI BOLD signal were made on a horizontal-bore Phillips 3 tesla scanner at Newcastle Magnetic Resonance Centre. Functional data were acquired using a single-shot gradient-recalled echo planar imaging sequence similar to that used to test the Rhesus macaques (volume acquisition time 2s; echo spacing = 9s; flip angle = 90^o; TE = 30ms; 28 slices, 4.6mm; in-plane field of view: 192 x $192mm^2$, on a grid of 64 x 64 voxels, with voxel resolution of 3 x 3 x $4.6mm^3$; inter-trial-interval = 9s; Figure 5.1). T1-weighted anatomical images were acquired in register with each functional scanning experiment using an MDEFT sequence with parameters TE: 4.6ms; echo spacing: 1300ms; 288 x 288 voxels; 150 slices with resolution 1.15 x 1.15 x $1.15mm^3$.

Human fMRI data analysis

First level GLM analyses with fixed effects (FEAT, FSL) were performed at the individual level. Unlike the Rhesus macaque experiment, it was unnecessary to combine several different imaging sessions in the same subject, therefore all data was registered to a standard template brain to allow comparisons between participants. These data were then combined using a cluster corrected ($p < 0.05$) higher-level analysis and projected to a surface-rendered standard template brain (FreeSurfer, McLaren *et al.*, 2009) for display purposes.

Human ROI analyses

Regions of interest for BA44/45 in both hemispheres were defined based on probabilistic maps (Harvard-Oxford Cortical Structural Atlas and the Juelich Histological Atlas, Eickhoff *et al.*, 2005; Desikan *et al.*, 2006). A vFC ROI was defined encompassing more ventral and medial parts of the inferior frontal gyrus, including the frontal operculum (Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008), and excluding areas BA44 and 45 (Figure 5.4). To ensure the results were comparable to the macaque data, the average BOLD activation in each voxel, across each participant was calculated. These data were normalised and analysed identically to the macaque data.

5.4.2. Results

Following habituation to exemplary consistent AG sequences, nine humans were scanned with fMRI while being presented with the consistent and violation testing sequences (Figure 3.1). FMRI BOLD signal to violation sequences relative to consistent test sequences revealed a number of highly significant clusters of activation (cluster corrected, $p < 0.05$; Figure 5.4A-D and Table 5.2). A large cluster of activation was observed, including the ventral portion of the inferior frontal gyrus, frontal opercular cortex and posterior regions of the frontal pole, in both hemispheres (Figure 5.4). These regions appeared to be ventral to Broca's area and its right hemisphere homologue (BA44/45). Strong clusters of activation were also observed in temporo-parietal areas, including the angular gyrus (BA39). This activation extended along the temporal lobe to include large parts of the posterior temporal lobe and middle temporal gyrus. Additional activation was also observed in the frontal pole and dorsal frontal cortex, excluding BA44 and BA45. Finally, subcortical activation was observed in the cingulate cortex. All of these clusters of activity appeared to be activated broadly bilaterally, and no strong left hemisphere bias was observed. To further investigate the role of BA44/45 and more ventral regions within the IFG, and to quantify any potential lateralisation effects in these regions, ROI analyses comparable to those used in macaques were performed.

ROI results

The human ROI analyses showed that, as seen in all of the macaques (Figure 5.3), the vFC region was significantly sensitive to violations of the AG (Bonferroni corrected one-sample *t*-tests: left: $t_{953} = 12.063$, $p < 0.001$; right: $t_{945} = 13.292$, $p < 0.001$; Figure 5.4E). As with the monkeys, this effect was not significantly lateralised (two-sample *t*-test between hemispheres: $t_{1898} = 0.930$, $p = 0.352$; Figure 5.4E). However, unlike the monkeys, BA44/45 in humans showed no significant sensitivity to violations of this AG (left: $t_{717} = 0.412$, $p = 1.0$; right: $t_{621} = 1.169$, $p = 1.0$; Figure 5.4F), and no differences were observed between hemispheres ($t_{1337} = 0.771$, $p = 0.441$). These results support previously reported human AGL fMRI results, suggesting that the processing of relatively simple FSGs strongly activate vFC regions, but that Broca's area (BA44/45) remains uninvolved (Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009).

Rhesus macaque and human comparisons

A direct comparison of the human and monkey ROI results was made using a repeated measures ANOVA with the factors: condition (consistent, violation), species (human, monkey), ROI (BA44/45, vFC) and hemisphere (left, right). This showed a significant overall sensitivity to violation vs. consistent AG sequences (main effect of condition: $F_{1,8129} = 206.1$, $p < 0.001$), with no significant interaction between condition and species ($F_{1,8129} = 3.54$, $p = 0.06$). A significant condition by ROI interaction was observed ($F_{1,8129} = 38.0$, $p < 0.001$) recapitulating that the vFC is more involved than BA44/45. Finally, there was an interaction between condition, ROI and species ($F_{1,8129} = 21.9$, $p < 0.001$) showing that BA44/45 is relatively more involved in the monkeys than in humans.

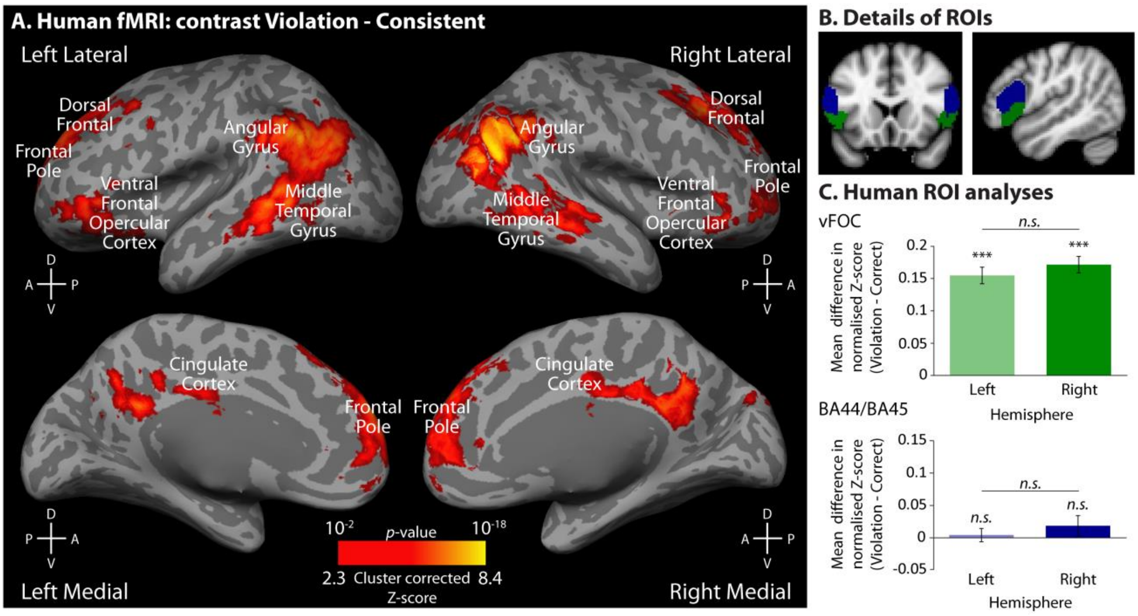


Figure 5.4 Human brain regions sensitive to AG violations and ROI analyses. A. Group statistical parametric map of sensitivity to AG violations (contrast: violation vs. consistent), all $p < 0.05$ cluster corrected. Results are displayed on rendered medial and lateral surface representations transformed to the MNI standard brain and using the Harvard-Oxford Cortical Structural Atlas and the Juelich Histological Atlas (Eickhoff *et al.*, 2005; Desikan *et al.*, 2006) human atlas to define anatomical regions (B). C. Normalised mean ROI voxel response differences (violation vs. consistent) in the fMRI signal for BA44/45 and vFC (see text) in the left and right hemisphere of the human brain. Abbreviations: D=Dorsal; V=Ventral; A=Anterior; P=Posterior; *** = $p < 0.001$.

Table 5.2 Anatomical locations of significant clusters in humans. MNI Coordinates and maximum *z*-scores for significantly activated clusters across 9 human participants.

Anatomical Location	**MNI coordinates**			**Cluster corrected max. *z*-score**	**Hemisphere**
	x	***y***	***z***		
Ventral Frontal Cortex Frontal Opercular Cortex, Frontal Orbital Cortex, Frontal Pole	48	26	18	4.94	Right
Ventral Inferior Frontal Cortex Frontal Opercular Cortex, Frontal Orbital Cortex, Frontal Pole	-52	38	-6	4.77	Left
Frontal Pole	28	28	44	6.49	Right
Angular Gyrus and Middle Temporal Gyrus	46	-66	28	9.11	Right
Angular Gyrus	-56	-52	42	7.37	Left
Middle Temporal Gyrus	62	-56	-18	4.79	Right
Cingulate Gyrus	-4	-50	32	5.96	Bilateral

5.4.3. Human conclusion

A number of perisylvian brain regions showed sensitivity to violations of this AG structure, including a number of areas comparable to those involved in natural language processing and those activated in Rhesus macaques. Ventral portions of the IFG, as well as temporal and temporo-parietal regions were sensitive to violations of the AG. Furthermore, significant increases in BOLD response were observed in the frontal pole and dorsal frontal regions, as well as in the cingulate cortex. However, both the GLM and ROI analyses suggest that BA44 and 45 were not involved in the processing of this AG in humans, raising interesting implications regarding the evolution of this region, and the functions of its structural homologues in nonhuman primates.

5.5. Discussion

These comparative fMRI experiments have shown that several perisylvian brain regions (including anatomically corresponding ventral frontal, temporal and parietal regions) in both humans and monkeys were sensitive to violations of the AG. These results demonstrate some similarities in how this AG structure is processed in the macaque and human brain, suggesting the functions supported by these regions may share a common evolutionary heritage. Furthermore, these correspondences suggest that Rhesus macaques may represent an appropriate animal model system in which to study aspects of language processing at a neuronal level. Finally, the patterns of activation observed in specific areas of the IFG suggest and constrain potential hypotheses regarding the evolutionary specialisation of these areas for language in humans.

Ventral portions of the IFG, including frontal opercular and insula areas (but not necessarily Broca's area or its presumed homologues, BA44 and BA45), were strongly activated in both species. The role of the IFG in language processing has been well documented (see Chapter 1, and e.g. Friederici, 2011). Activation in this region is seen to increase with the syntactic complexity of sentences (e.g. Constable *et al.*, 2004; Bornkessel *et al.*, 2005). Furthermore, both auditory and visual AGL paradigms have consistently reported activation in this region suggesting that ventral parts of the frontal cortex might be critically engaged in evaluating the structure of sequences, both in natural language sentences and in artificial grammars (Petersson *et al.*, 2004; Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008; Udden *et al.*, 2008; Bahlmann *et al.*, 2009; de Vries *et al.*, 2010; Petersson *et al.*, 2012). Therefore, results showing that this area is activated, by the same AG structure, in both human participants and nonhuman primates, provide valuable evidence the processes supported by this region may be generic and evolutionarily conserved, rather than language specific and uniquely human.

Correspondence between the species was also observed in temporo-parietal regions, Brodmann area 7 in macaques and the angular gyrus (BA39) in humans, both of which lie below the intra-parietal sulcus. There has been considerable uncertainty regarding whether these regions might be functionally homologous (Karnath *et al.*, 2001), given that human BA7 lies above the intra-parietal sulcus. These results suggest that BA7 in macaques and BA39 in humans share a functional correspondence that involves evaluating the structural relationships between elements in a sequence. In humans, the posterior temporal lobes and the angular gyrus are involved in both semantic and syntactic processing (Kinno *et al.*, 2008;

Friederici *et al.*, 2010), and appear to be particularly strongly activated by the integration of these two types of information (Friederici *et al.*, 2003; Friederici, 2011). However, AGL paradigms have also produced activation in these regions (Friederici *et al.*, 2006a; Bahlmann *et al.*, 2009), suggesting that this area is involved in processing the structure of sequences of elements, even in the absence of semantic content. It is interesting to note that the angular gyrus in humans forms part of the dorsal pathway (e.g., Friederici *et al.*, 2006a; Hickok & Poeppel, 2007; Saur *et al.*, 2008), normally associated with evaluating nonadjacent relationships or relatively complex AG or syntactic processing. In nonhuman primates this pathway is less evident than the ventral pathway (Rilling *et al.*, 2008), and no nonhuman primates have demonstrated the ability to learn AG structures that would activate this pathway in humans (see Chapter 1, and e.g., Fitch & Hauser, 2004). Therefore, these results present the interesting possibility that not only the ventral but possibly also aspects of the dorsal pathways may be engaged in the processing of simpler AGs in nonhuman primates. While DTI tractography was not possible in this experiment, detailed investigation into the connectivity between temporo-parietal regions and the IFG in nonhuman primates might provide additional insights into the connectivity of the network involved in AGL, and how this relates to the language or AGL network in humans.

While the correspondences observed between the species are notable, some differences in the areas activated between the species were in evidence. For example, humans showed sensitivity to violations of the AG in the frontal pole and dorsal frontal cortex, which were not significantly involved in the macaques. These regions have been shown to be involved in decision making in humans (e.g. Bechara *et al.*, 2000; Bechara & Van der Linden, 2005). While, efforts were taken to minimise the differences in how the humans and monkeys were studied, the differences between human participants and nonhuman primates necessitated some experimental differences (e.g., AG learning evaluated with implicit measures in monkeys and with explicit responses in humans). The main objective of these experiments was to use the same stimuli and AG paradigm with the humans and monkeys, however, some of these task differences could have contributed to the lack of clear cross species correspondence in the dorsal frontal, frontal pole or subcortical regions. Differences were also observed in the anterior temporal lobe, which is part of the ventral pathway, and is known to be involved in language processing in humans, particularly at the phrase level (e.g. Friederici *et al.*, 2000; Hickok & Poeppel, 2007; Friederici, 2011). While this area was recruited in the macaques, no significant increase in activation was observed in this area in the human participants. Conversely, the human results demonstrated significant activation in middle temporal lobe, including the middle temporal gyrus. This region has been implicated in speech processes bilaterally in humans (Hickok & Poeppel, 2007) but is a gyrus that is not present in macaques. The results of these fMRI experiments demonstrate some interesting inter-species differences. However, this might be expected, given that humans and Rhesus macaques last shared a common ancestor approximately 30 million years ago (Steiper & Young, 2006). Human language has clearly evolved enormously in complexity and scale relative to the communication systems of all other primates, and it is unsurprising that this might lead to differences in how AG structures, designed to emulate aspects of language, might be processed. Nevertheless, correspondences in some key brain areas, known to be involved in language processing in humans, suggest that this AGL task recruits some comparable areas, which might support more generic, conserved functions in the human brain.

Beyond the results of the GLM analyses, the ROI analyses provide further insights into the specific function of different IFG regions. Humans have been shown to rely on evolutionarily older parts of the ventral frontal-opercular cortex (Sanides, 1962; Friederici *et al.*, 2006a), not including Broca's area (BA44/45), to evaluate adjacent relationships or simple AG structures (Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009), as part of an initial syntax-building function in the brain network for language (Friederici, 2011). In the human participants and all three macaques tested, vFC regions, ventral to BA44 and BA45, were significantly activated by violations of the AG. These results support the hypothesis that the vFC underpins evolutionarily conserved, generic functions for assessing the relationships between elements in this relatively simple AG structure.

By contrast, Broca's area represents a more recent evolutionary specialisation involved only in more complex aspects of sequence processing (e.g. Petersson *et al.*, 2004; Friederici *et al.*, 2006a; Friederici, 2011). In the human participants, this region showed no activation to AG violations in either hemisphere, suggesting that the AG structure used here was insufficiently complex to recruit Broca's area. By contrast, significant activation in BA44 and BA45 was observed in two of the Rhesus macaques. A number of hypotheses have been proposed regarding the potential existence of a functional homologue of Broca's area in nonhuman primates, and the implications this might have on nonhuman AGL (see Introduction and, Friederici, 2004; Friederici *et al.*, 2006a; Petkov & Wilson, 2012). However, the results of these experiments suggest that while BA44/45 might represent a structural homologue of Broca's area based on cytoarchitectonic similarities (Brodmann, 1909), these regions appear to have functionally differentiated in humans. Broca's area in humans represents a unique evolutionary specialisation for processing complex linguistic relationships (Petersson *et al.*, 2004; Friederici *et al.*, 2006a; Friederici, 2011). However, these results suggest that the comparable area in Rhesus macaques is, to some extent, activated by simpler AG structures. Nonhuman primates have previously been reported to learn simple, but not more complex AG structures that produce activation in Broca's area in humans (Fitch & Hauser, 2004). Therefore, these results raise interesting questions, regarding whether the upper limits of nonhuman animal AGL capabilities might be imposed by this apparent inter-species difference. Alternatively, if nonhuman primates were to be shown to be able to learn AGs that produce activation in Broca's area in humans (possibly by explicit training rather than implicit habituation), what patterns of activation might we observe in BA44/45 and elsewhere? The results of these fMRI experiments are insufficient to delineate the evolutionary history of the IFG, and the language related abilities that it supports, but they do suggest that while the function of the vFC appears to be relatively conserved in both species, BA44/BA45 might have shown more divergence between species.

The human language network is known to be left lateralised (e.g., Friederici & Alter, 2004; Hickok & Poeppel, 2007; Tyler & Marslen-Wilson, 2008; Friederici, 2011). Although activation is reported to some extent in both hemispheres in language related tasks, these effects are typically much stronger in the left hemisphere. The fMRI results presented here show no such lateralisation effect, in either the GLM or ROI analyses, in either species. The human fMRI results reveal a clear bilateral distribution of activation. In the macaques, two animals showed bilateral fronto-temporal activation, based on statistical comparisons between the hemispheres, and one monkey showed increased activation in the right hemisphere, at least in the vFC. These data are insufficient to conclude that finite state AGs are primarily supported by a right lateralised network of brain areas in Rhesus macaques. Instead, these results, along with the data from human participants, suggest that rather than recruiting left hemisphere brain regions, this artificial grammar appears to be broadly bilaterally processed.

This result might not be unexpected in nonhuman primates. If the lateralisation effects observed in the language network represent an adaptation to specific linguistic demands (Friederici & Alter, 2004), then it is unlikely that such lateralisation would also be present in non-linguistic primates (Wilson & Petkov, 2011). In the human participants this result is more surprising. Previous AGL studies have reported activation only in the left IFG, even using only simple AGL paradigms (Friederici *et al.*, 2006a; Bahlmann *et al.*, 2008).

However, in these studies, unlike those presented here, quantitative analyses of lateralisation effects were typically not reported, therefore it is possible that activation may have occurred bilaterally, but only reached the significance threshold in the left hemisphere. Alternatively, if the lateralisation of the language network represents a specialisation for certain features of language, it is possible that the AG structure used in these experiments is simple enough to be processed bilaterally, and does not require more specialised processing in one particular hemisphere.

Indeed, the results of these experiments may provide the first direct comparative evidence in support of a 'dual neurobiological language systems' hypothesis (Bozic, *et al*, 2010; Wright *et al.*, 2012; Bozic, *et al*, 2013; Marslen-Wilson, *et al.*, in press), and may also extend it in important ways. This hypothesis proposes that in modern humans, specialisations for core syntactic language functions depend on a left-lateralised fronto-temporal system, and that this left-lateralised system is functionally integrated with a more ancestral, bilaterally distributed system. The bilateral system is suggested to support sound to meaning mapping, as well as aspects of semantic and pragmatic interpretation (Bozic, *et al*, 2010; Marslen-Wilson, in press). Given that the current results are based on finite-state AG learning, these findings suggest that the bi-hemispheric system also supports relatively simple sequence learning in both humans and monkeys. These results raise the possibility that language processes in modern humans are functionally integrated with an ancestral system that may have evolved from domain-general, non-linguistic cognitive processes seen in our extant primate relatives.

5.6. Conclusion

The results of these experiments provide important, novel evidence suggesting that AG structures appear to be processed by a broadly comparable perisylvian network of frontal, temporal and parietal brain regions in both human participants and Rhesus macaques. This suggests that while human language is undoubtedly unique, some of the brain areas that support the processing of some features of language appear to be evolutionarily conserved and therefore likely generic in function, suggesting that the Rhesus macaque might represent a valuable animal model system. However, while a number of regions showed similarities across the species, differences were observed, particularly between Broca's area (BA44/45) and its structural homologue in nonhuman primates. These results suggests that some uniquely human language abilities, absent in nonhuman primates, might be supported by BA44/45 in the IFG and potentially also by the left lateralisation of the human language network, which appear to be absent in nonhuman primates. These experiments provide valuable insights into the evolution of the brain regions that support language in humans. Furthermore, they provide the first neuroscientific evidence that Rhesus macaques may be a valuable model species in which to study these sequence structure learning abilities at a neuronal level, and have dramatic implications for the further study of language related abilities in nonhuman primates.

Chapter 6: General Discussion

Neuroimaging research is clarifying the roles of the human brain areas that support the processing of language and of artificial grammars. Concurrently, behavioural AGL paradigms have demonstrated that some abilities, which are potentially related to aspects of human language, appear to be present in a range of nonhuman animals. The experiments described here sought to combine these research strands to investigate the AGL capabilities of nonhuman primates, and the brain areas that support them. In this thesis, I have presented novel methods for comparing AGL capabilities across species and studies, and for more precisely and objectively assessing these abilities in nonhuman primates. I have demonstrated that both Rhesus macaques and common marmosets are able to learn an artificial grammar that emulates some important features of language syntax, but that only macaques show learning that cannot be explained by simple strategies. Finally, comparative fMRI experiments have shown that AGL produces activation in a range of comparable of perisylvian brain areas in both macaques and humans, suggesting that some of the mechanisms and brain areas that support aspects of language in humans are supported by generic rather than human- or language-specific systems.

Beyond these results, this research had two long term goals. Firstly, to provide insights into how certain aspects of language may have evolved, and secondly to begin the development of potential animal model systems in which any abilities shared between species could be investigated at a neuronal level. Our understanding of the cognitive mechanisms and network of brain areas underpinning language learning and processing are increasingly well understood (e.g. Friederici, 2002; Vigneau *et al.*, 2006; Hickok & Poeppel, 2007; Tyler & Marslen-Wilson, 2008; Friederici, 2011). However, the evolutionary origins of language, which is not only a uniquely human trait but may also be a defining characteristic of our species and a key factor in our evolution, remain unclear (Christiansen & Kirby, 2003b; Fitch, 2010; Hurford, 2012). As a result of this lack of clarity regarding the origins of human language abilities, it has previously been difficult to imagine how any non-linguistic species might inform us about the neural mechanisms supporting language in the human brain. However, if such an animal model system were to be developed, it would have the potential to provide invaluable insights into how certain features of language may be processed at a neuronal level. In this chapter, I will discuss a number of key obstacles faced while attempting to reach these goals, how the research presented in this thesis has attempted to overcome these challenges to clarify the path ahead, and what questions and additional research remain if these goals are to be achieved.

6.1. Comparing different AGL capabilities

Artificial grammars are designed to emulate specific features of language, which can be studied independently of other features that nonhuman animals may be unable to learn (Reber, 1967). Therefore, by design, no AG structure aims to represent the complexity of natural language. Nonhuman animals have been tested with a wide variety of different AGL paradigms designed to investigate different language related abilities (Fitch & Hauser, 2004; Gentner *et al.*, 2006; Murphy *et al.*, 2008; Saffran *et al.*, 2008; Hauser & Glynn, 2009; van Heijningen *et al.*, 2009; Abe & Watanabe, 2011; Stobbe *et al.*, 2012). However, AGL is a not a single capability, the abilities required to recognise violations of a specific AG vary with the structure of that AG. Therefore, the ability to learn one structure need not necessarily generalise to others. For example, an animal may be able to categorise stimuli into A and B classes, and learn that the $(AB)^n$ structure produces the invariant ABAB sequence (Fitch & Hauser, 2004). However, based on this result, it might be unreasonable to conclude that this species would also be able to learn a more varied, AG such as the one used in these experiments, or more complex variants such as that of Reber (1967). Of course, the best way to assess a species' ability to learn or process a feature of language would be to test the animals using an AG emulating that specific feature. However, language is very complex, and a huge variety of different AGL paradigms could be developed before all potentially language related abilities have been assessed. AGL experiments in nonhuman animals are time consuming endeavours. Therefore, it is not possible to test nonhuman animals on a range of AGs wide enough to model all the different abilities that might be required for human language. A method to compare different AG structures, and the abilities that support them, may present a more efficient way of exploring the AGL abilities in nonhuman animals and identifying shared abilities between species.

The Chomsky Hierarchy (Chomsky, 1957) and more recent variants (Berwick *et al.*, 2011; Jaeger & Rogers, 2012) provide mathematical descriptions of grammars of different levels of generative power. However, as discussed in Chapter 2, distinctions between relatively simple, finite-state grammars and more complex, context-free grammars might be insufficient to inform us about the extent of nonhuman animals' abilities (Hurford, 2012; Jaeger & Rogers, 2012; Petkov & Wilson, 2012), since there is currently no evidence that animals can learn anything more complex than a finite-state grammar (Perruchet & Rey, 2005; van Heijningen *et al.*, 2009; Beckers *et al.*, 2012; Berwick *et al.*, 2012; ten Cate & Okanoya, 2012). This has prompted several authors to highlight the need for alternative methods of quantifying the complexity of simpler AG structures (de Vries *et al.*, 2011; Hurford, 2012; Jaeger & Rogers, 2012; Petkov & Wilson, 2012), to allow the abilities of nonhuman animals to be more meaningfully compared across AGL paradigms. The quantitative parameter space proposed in Chapter 2 represents one such method (Figure 2.3 and Figure 2.4).

Different AG structures aim to model different features of language. However, the core of all AGL paradigms involves learning the structure governing sequences of stimuli. Two key dimensions on which all AGs can be compared are the number of elements or stimulus classes which contribute to the structure, and the predictability or linearity of the structure based on the number of different legal transitions between elements. These two dimensions allow us to plot a measure of the complexity of different AG structures onto a common parameter space where they can be objectively compared (see Chapter 2 and Figure 2.3 and Figure 2.4). This allows not only the direct comparison of a species' abilities to learn different AG structures, but also facilitates predictions regarding an animal's ability to learn previously untested structures. For example, if an animal were able to learn a number of AG structures of a certain level of complexity, it is likely that if all other factors and additional demands are kept consistent, they would also be able to learn any simpler AGs. To objectively draw such comparisons between AG structures, a quantitative measure of complexity is required. Furthermore, data regarding a species' ability to learn a number of different AGs allow us to map out an area of the parameter space representing the structures that species is able to learn. This could be used to guide hypotheses regarding future AGL experiments in that species, and to draw comparisons between species. Finally, plotting the abilities of different species on a parameter space such as the one presented in Chapter 2, helps to clarify gaps in our understanding about the AGL capabilities of nonhuman animals. This prompted the selection of an AG that has received relatively little investigation and represented the opportunity to provide valuable insights into language related abilities in nonhuman primates.

The non-deterministic, AG structure used in these experiments was developed from Saffran, et al (2008). Unlike many of the AG structures previously used to test nonhuman animals (Fitch & Hauser, 2004; Gentner *et al.*, 2006; Murphy *et al.*, 2008; Hauser & Glynn, 2009; van Heijningen *et al.*, 2009; Stobbe *et al.*, 2012), this AG departs from the requirement that participants initially categorise stimuli into one of two classes, and then learn a simple, invariant structure (e.g. ABAB generated from $(AB)^n$; Fitch & Hauser, 2004; see Section 1.3.1). All species tested have been shown to learn at least the simpler AGs of this type, e.g. $(AB)^n$. Furthermore, despite the intent of some studies (Fitch & Hauser, 2004; Gentner *et al.*, 2006), even the more complex structures (i.e. A^nB^n) tested fail to model the hierarchical, centre-embedding that represents context-free grammars (Perruchet & Rey, 2005; van Heijningen *et al.*, 2009). Therefore, it is possible that the failure of nonhuman primates to learn the more complex A^nB^n structure (Fitch & Hauser, 2004), might be due to its increased nonlinearity relative to the $(AB)^n$ structure. Therefore, the AG structure used in this thesis presents an opportunity to test whether nonhuman primates are able to learn a more nonlinear AG structure.

The AG used in these experiments does not require categorisation of the stimuli; instead participants are required to extract statistical regularities or rules regarding the relationships between 5 different nonsense word elements from a much wider variety of different habituation sequences. Two previous studies have attempted to test similar AGL paradigms in nonhuman animals (Saffran *et al.*, 2008; Abe & Watanabe, 2011). However, experimental difficulties have made the interpretation of these results challenging (see Chapter 2 and the following section, Beckers *et al.*, 2012; Berwick *et al.*, 2012). While this structure, like all AGs, is insufficient to inform us about all of the different processes required for human language learning, it does present a valuable opportunity to assess whether nonhuman primates are able to learn a relatively complex, non-linear AG (see Chapter 2 and Figure 2.3), and whether they can extract statistical rules from a varied input rather than simply noticing sequences which vary from a fixed pattern (e.g. Fitch & Hauser, 2004).

An objective method by which to quantify the complexity of different artificial grammars is an important tool. The parameter space presented in Figure 2.3 represents one such tool, however, as discussed in Chapter 2, additional factors may need to be included. For example, the calculations of linearity used in this model assume that only local relationships between adjacent elements (represented by 1st order Markov processes, Hurford, 2012) are employed in learning the structures. Furthermore, this parameter space plots the position an AG would occupy if it was learned in its entirety, and participants did not base their responses on simpler strategies. However, if rather than learning that the A^nB^n structure produces the sequence AABB, animals simply responded to the repetition of the A or B element (van Heijningen *et al.*, 2009; and see Chapter 2) then the animal cannot be said to have learned the complete structure, only a small part of it, and therefore its position in the parameter space may be inaccurate. Therefore, in order to accurately interpret or compare the AGL capabilities of nonhuman animals both their abilities to learn the AG structure, as well as the strategies used, must first be rigorously and accurately assessed. In the next section I will discuss how the methodological developments presented here begin to overcome some of these problems.

6.2. Assessing AGL abilities in nonhuman primates

If any nonhuman animal is to inform us about language evolution, or to act as an animal model system, it is necessary to rigorously confirm that the animal is able to learn a given artificial grammar. While the production of larger dishabituation responses to sequences which violate an AG, relative to those that are consistent with the structure does suggest some level of learning, these results may be insufficient to determine the strategies upon which their learning may be based. For example, if an animal were to be habituated to sequences of the form AABB, generated by the A^nB^n structure, and then be tested with both AABB and ABAB sequences (from the $(AB)^n$ structure, Fitch & Hauser, 2004) dishabituation responses could be interpreted in a number of ways. The animal may have learned that the sequences must be made up of a number of centrally embedded 'AB' phrases (i.e. A[AB]B). However it is also possible, and likely more parsimonious, to attribute these responses to, for example, the absence of the repeated 'AA' or 'BB' elements, or the alternating structure of the ABAB sequences (Perruchet & Rey, 2005; Gentner *et al.*, 2006; Corballis, 2007; van Heijningen *et al.*, 2009). Even with more complex testing sequences, such as AABB vs. ABBB (Gentner *et al.*, 2006), simpler strategies, such as recognising an imbalance of A and B elements, are still possible (van Heijningen *et al.*, 2009). Similarly, when Saffran and colleagues presented cotton-top tamarins with an AG similar to the one used here, they tested the animals with the same consistent, grammatical sequences to which the animals were habituated (Saffran *et al.*, 2008). Therefore, it is impossible to rule out the possibility that the tamarins simply responded more strongly to novel sequences than familiar ones. Bengalese finches have also been tested with a similar AG (Abe & Watanabe, 2011), however in these experiments acoustical differences between consistent and violation sequences made it difficult to attribute responses to genuine AGL (Berwick *et al.*, 2011; Beckers *et al.*, 2012). These experimental design problems make it difficult to determine the extent to which different animals (including humans) may have learned an artificial grammar, and therefore represent a critical challenge in designing appropriate AGL experiments.

The accurate assessment of nonhuman animals' AGL capabilities requires appropriately designed experiments. In the experiments described in this thesis, care was taken to include test sequences that might help to identify the learning strategies animals used (see Chapter 3), which might identify potential differences between species. The inclusion of 'familiar' and 'novel' consistent sequences and violation sequences that either 'begin with A' or 'do not begin with A' provide considerably more information than experiments with only two test conditions (e.g. Fitch & Hauser, 2004; Saffran *et al.*, 2008). The inclusion of these conditions allow experiments that not only aim to demonstrate that animals are able to recognise AG violations, but also to determine what strategies their responses might be based on. Erroneous conclusions, based on the over-interpretation of results, have the potential to limit rather than further our search for precursors to aspects of language in nonhuman animals and the development of animal model systems.

While the inclusion of additional conditions to these experiments provides an added level of rigour relative to a number of previous studies, these do not overcome all the potential problems associated with traditional AGL experiments. Firstly, all previous nonhuman primate (although not songbird) AGL experiments have been carried out by the same laboratory and all relied on a video-coding paradigm similar to the one described in Chapters 3 and 4 (Hauser *et al.*, 2001; Fitch & Hauser, 2004; Newport *et al.*, 2004; Saffran *et al.*, 2008; Hauser & Glynn, 2009). In the experiments in Chapters 3 and 4, every effort was made to refine this approach to make it as objective as possible. However, any analysis based on the ratings of human observers carries with it an inherent level of subjectivity. Furthermore, the experiments presented here collected data not only on whether or not the animals responded to a stimulus (frequency of responses), but also the duration and strength of these responses (see Chapters 3 and 4). This additional information provides some interesting insights; however, these measures still have insufficient temporal detail to measure effects within individual testing sequences. The eye-tracking paradigm presented in Chapters 3 provides both an objective measure of responses and sufficient temporal precision, which helps to overcome both of these problems.

While care was taken to design the testing sequences of this experiment to provide as much insight as possible into the AGL capabilities of nonhuman primates, certain simple strategies could still produce a pattern of responses similar to more complex AGL. No differences in the macaques' responses were observed between violation sequences based on whether or not they contained a violation in the first sequences position (whether they began with A or not). Therefore, the animals could not have relied solely on the initial element to guide their responses. However, in an attempt to balance the test sequences on a number of other dimensions, both of the violation test sequences that 'begin with A' contained the (violation) transition from A to F in their second position (Figure 3.1). Therefore, it is possible that the animals responded to this relatively simple transition. Traditional AGL paradigms would be unable to rule out this possibility, however the temporal precision of the eye-tracking approach permitted analyses of response to specific elements in the test sequence, showing that (in at least two out of the three animals tested), that the responses could not be attributed only to this relatively simple rule (Figure 3.8). The developments presented in these experiments present potential methods to more objectively assess not only whether nonhuman animals are able to recognise violations of an AG, but also to explore the strategies or cues that might underpin their AGL capabilities.

6.3. Insights into nonhuman primate AGL capabilities

The methods and techniques presented in this thesis represent an important increase in the level of specificity that can be attained in AGL experiments, and therefore provide valuable tools in assessing the sequence-structure learning capabilities of nonhuman animals. The results obtained in these experiments provide novel insights into the AGL capabilities of nonhuman primates. Both Rhesus macaques and common marmosets were more responsive to violation sequences than consistent sequences. Furthermore, both the video-coding and eye-tracking experiments in macaques revealed no differences in responses between 'familiar' and 'novel' sequences or those that 'begin with A' and 'do not begin with A'. The macaques also responded significantly more frequently and for longer durations to violation sequences that 'begin with A' than 'novel' consistent sequences (see Chapter 3). These results provide important evidence, which could not be obtained from previous nonhuman primate AGL studies (e.g., Fitch & Hauser, 2004; Saffran *et al.*, 2008), that Rhesus macaques do not appear to respond based on the familiarity of test sequences or violations in the initial sequence position. By contrast, the video-coding experiment in common marmosets showed no differences in the frequency or strengths of responses to violation sequences relative to consistent ones. Furthermore, even though a significant difference was observed in the marmosets' response durations, these results could be explained by simple strategies, suggesting that marmosets' AGL may have been relatively shallow compared to that of macaques. These results demonstrate a previously unreported level of AGL in nonhuman primates, and suggest a potential evolutionary gradient, with more closely related Old World monkeys possessing more complex AGL capabilities than New World monkeys.

Evidence that any nonhuman primate can learn this AG suggests that the abilities which underpin the detection of violations of the AG structure are not unique to humans or specific to language. Furthermore, confirmation of AGL in a second species, albeit to a simpler level than macaques, reduces the likelihood that such abilities may have arisen in humans and nonhuman primates by convergent evolution rather than common descent, and provides evidence for the hypothesis that they might be supported by similar mechanisms across species. Data demonstrating that more closely related Old World monkeys appear to learn more about the AG structure provide interesting insights regarding how and when these abilities might have evolved, and suggests that Rhesus macaques might represent a more suitable candidate model system. However, these behavioural data in two species are insufficient either to identify what natural nonhuman primate abilities might support AGL in non-linguistic species, or to determine whether a species might function as a useful neurobiological animal model system. Behavioural correspondences need not represent the same cognitive mechanisms or produce activation in homologous brain regions across species. Comparative neuroimaging has the potential to identify whether the abilities observed in two species are supported by the corresponding areas of the brain.

6.4. Exploring the brain areas supporting AGL in primates

Behavioural correspondences alone represent insufficient information to confirm that the cognitive mechanisms and brain areas that support an ability are consistent across species. Without evidence demonstrating functional homologies in the brains of human and nonhuman primates, it is possible that the language network in humans represents an entirely unique evolutionary adaptation, and that primate AGL is supported by wholly different mechanisms. While this is true of any behavioural correspondence between two species, it might be particularly relevant when discussing capabilities relating to natural language, which is not shared by any other species. In order to conclude that the behavioural similarities observed in humans and nonhuman primates (see Chapter 3 and 4) might be supported by homologous brain regions, and therefore potentially comparable cognitive mechanisms, comparative neuroimaging was required. Furthermore, activation in response to AG violations in regions known to be involved in human language processing, might additionally suggest that AGs are processed similarly to aspects of language in humans, and therefore do not simply represent a generic pattern recognition ability independent of language.

Language processing in the human brain is supported by a number of perisylvian brain areas in the left hemisphere, including the inferior frontal gyrus, temporal lobe and temporo-parietal regions (Broca, 1861a; Damasio & Geschwind, 1984; Binder *et al.*, 1997; Catani *et al.*, 2005; Hickok & Poeppel, 2007). Both natural language and AGL fMRI experiments have particularly emphasised the involvement of IFG in processing the syntax of language or the structure of AG sequences (Petersson *et al.*, 2004; Friederici *et al.*, 2006a; Hickok & Poeppel, 2007; Tyler & Marslen-Wilson, 2008; Bahlmann *et al.*, 2009; Makuuchi *et al.*, 2009; Petersson *et al.*, 2012). Specifically, these studies found that ventral frontal cortex regions (vFC) are recruited in the processing of all syntax-related tasks and AGL paradigms, regardless of complexity (Friederici *et al.*, 2006a; Bahlmann *et al.*, 2009). However, Broca's area (BA44/45, located dorsal to the vFC) is only activated by more complex structures (Petersson *et al.*, 2004; Friederici *et al.*, 2006a; Bahlmann *et al.*, 2009; Makuuchi *et al.*, 2009; Petersson *et al.*, 2012). Therefore, if the AGL capabilities observed in Chapters 3 and 4 were supported by similar cognitive mechanisms in humans and monkeys, it is likely that we would see some correspondences in these areas. Conversely, if human language were to represent a truly unique specialisation, absent in all nonhuman species, then the behavioural AGL results observed would be supported by different cognitive processes which we might expect to be supported by different neural substrates.

The fMRI experiments in humans and Rhesus macaques presented in Chapter 5 showed that in both species violations of the AG produced activation in ventral portions of the inferior frontal gyrus (particularly vFC), temporal and temporo-parietal regions. Some inter-species differences were observed in the temporal lobe, dorsal-frontal and frontal pole regions; however, the general pattern of activation suggested that a number of comparable brain areas were activated in both species. These broad correspondences suggest that AGL capabilities and the brain areas that support them likely share a common evolutionary history, having evolved from the same regions in a shared ancestor. Furthermore, evidence that activation is observed in comparable regions in both species, and that in humans these regions are involved in the processing of natural language, is the necessary first step to developing an animal model system in which to study aspects of language with techniques which are not typically possible in humans.

In addition to these findings, region of interest analyses were performed to investigate patterns of activation in the IFG in more detail. In both species strong activation was observed in ventral parts of the frontal cortex (vFC), suggesting that this is a key, conserved region in the processing of type of adjacent relationships present in this AG structure. However, differences were observed in BA44/45, dorsal to these vFC regions. Humans showed no activation to this AG in Broca's area, or its right hemisphere homologue (BA44/45). This result supports a number of previous studies implicating Broca's area only in the processing of more complex, non-adjacent or hierarchical features of syntax or AGL (Petersson *et al.*, 2004; Friederici *et al.*, 2006a). However, Rhesus macaques did appear to show some activation in this area. These results suggest that while BA44/45 in the macaques might be a structural homologue of Broca's area in humans, there has been some functional differentiation. It appears that Broca's area in humans may represent an evolved specialisation for complex language processing, while in the Rhesus macaques tested here, and potentially other nonhuman primates, BA44/45 is involved in processing simpler structures. These results might support those of previous behavioural work which has suggested that nonhuman primates (cotton-top tamarins) were unable to learn AGs of the level of complexity that produces activation in BA44/45 in humans (Fitch & Hauser, 2004; Friederici *et al.*, 2006a). It is therefore possible that evolutionary specialisation of Broca's area (in combination with other language adaptations) might be required for processing more complex features of language, or those associated with certain artificial grammars. However, activation in Broca's area has been reported in humans following habituation to the A^nB^n AG structure (Friederici *et al.*, 2006a), in this case, with no requirement that the participants learned specific associations between A and B elements (as in Bahlmann *et al.*, 2008; Bahlmann *et al.*, 2009). This experiment only required participants to discriminate between AABB and ABAB sequences, and yet still produced activation in Broca's area (Friederici *et al.*, 2006a). Rhesus macaques were able to learn the current AG structure implicitly, only through passive exposure. Therefore, whether they might be able to learn AGs such as the A^nB^n structure, perhaps with explicit training, remains an empirical question. If this were the case, fMRI data regarding the brain areas used to process this structure could be highly informative. The extent to which different brain areas involved in the processing of human language might have specialised to support language since our last common ancestor with macaques remains an open question. However, the results of these comparative fMRI experiments demonstrate that important similarities can be observed in the brains of both species in response to a relatively simple AG, designed to emulate some of the variability of the structure of sentences in natural language.

The results of these experiments have important implications not only for our understanding of the evolution of language, but also for the development of nonhuman primates as animal model systems in which to investigate aspects of language processing at a neuronal level. Correspondences observed between the brain regions activated by the same AG structure in humans and macaques suggest that the language network in humans likely evolved from brain regions shared with our last common ancestor. Therefore the same neuronal mechanisms might support comparable cognitive mechanisms across species, suggesting that the Rhesus macaque, and potentially other nonhuman primate species, may function as a valuable animal model. Electrophysiological recordings, either from single or multiple neurons or via local field potentials (e.g. Perrodin *et al.*, 2011) can be used to target areas localised by fMRI experiments (Petkov *et al.*, 2008b) such as those presented here. These techniques have the potential to investigate the mechanisms supporting these capabilities at a level of detail and temporal precision that is impossible using non-invasive imaging in humans, and in locations inaccessible to depth electrodes involved in the treatment of neurological problems, including epilepsy (e.g. Kumar, *et al.*, 2011). Furthermore, cortical manipulations including micro-stimulation (Petkov *et al.*, 2008c) or reversible inactivation of specific brain areas (Bartolo *et al.*, 2009) present additional possibilities to investigate how these regions contribute to AGL, and how they interact with other, associated regions.

Strong general correspondences were observed between macaques and humans, suggesting some areas, including ventral portions of the frontal cortex, are critically involved in processing this non-deterministic AG structure. However, there were also some differences between how the brains of macaques and humans responded to the AG. These results suggest that while some functions might be supported by comparable brain areas, other regions, including Broca’s area, may have specialised and now support different, potentially unique, functions in humans. Beyond providing interesting insights into the evolution of these areas, and of language processing in humans, such results might constrain the range of language related abilities for which nonhuman primates might be an appropriate animal model.

Human language is unique and an order of magnitude more complex than any abilities possessed by nonhuman primates. Therefore, while these experiments have demonstrated key correspondences in AGL capabilities and the brain areas that support them, some aspects of language are likely to be too complex for any nonhuman primate to learn. Therefore, it is not only important to consider the areas of correspondence, which can provide valuable insights into the evolution of these abilities and potentially about the neuronal mechanisms that support them, but also where human language might be unique, and the human brain might have specialised beyond the maximum capabilities of nonhuman animals. However, the testing of additional AG structures might help clarify the abilities shared between humans and nonhuman animals, and highlight both the features of language, and the brain areas supporting them, for which the human brain is particularly specialised and which might be conserved in nonhuman primates. Experiments in other nonhuman species might provide further insights into the language evolution, and the potential to develop additional animal model systems. Finally, additional techniques, including the neurophysiological approaches that are not possible in humans, in conjunction with this and future imaging work, have the potential to inform us about language related processes in a level of detail that has previously been impossible.

6.5. Conclusion

The goals of this research were to move toward a better understanding of the evolution of language and the development of animal model systems, which might allow us to explore the neuronal mechanisms supporting language related processes. The experiments reported in this thesis demonstrate that nonhuman primates are able to learn a non-deterministic AG, which emulates aspects of the variability present in the syntactic structure of natural language. Furthermore, at least in macaques, these abilities cannot be attributed to simple strategies. Finally, in both macaques and humans comparable perisylvian brain regions were sensitive to violations of this AG structure. These results suggest that while the brains of humans might be uniquely specialised for language, some of the processes underlying language processing might represent generic mechanisms supported by brain areas that share a common evolutionary heritage with those observed in extant nonhuman primates. These results allow the development of nonhuman animal model systems, which offer the potential to inform us about language processing at a neuronal level, which is not possible in humans.

Appendix: Other work, published or in preparation, by the author

7.1. Journal articles

Wilson, B., Kikuchi, Y., Sun, L., Hunter, D., Dick, F., Smith, K., Griffiths, T., Marslen-Wilson, W. D., Petkov, C. I. (submitted). Artificial-grammar learning engages evolutionarily conserved regions of frontal cortex in humans and macaques.

Wilson, B., Slater, H., Kikuchi, Y., Milne, A. E., Marslen-Wilson, W. D., Smith, K. & Petkov, C. I. (2013), Auditory artificial grammar learning in macaque and marmoset monkeys. *Journal of Neuroscience*, **33**, 48, 18825-18835.

Petkov, C. I. & **Wilson, B**. (2012). On the pursuit of the brain network for proto-syntactic learning in nonhuman primates: Conceptual issues and neurobiological hypotheses. *Philosophical Transactions of the Royal Society B,* **367**, 2077-2088.

Wilson, B. & Petkov, C. I. (2011). Communication and the primate brain: Insights from neuroimaging studies in humans, chimpanzees and macaques. *Human Biology*, **82**, 2, 153-173.

Petkov, C. I. & **Wilson, B.** (2011). Functional imaging of brain regions sensitive to communication sounds in primates. *Interspeech 2010*, 2494-2497.

7.2. Poster presentations

Wilson, B., Collison, M. G., Slater, H., Hunter, D. M., Smith, K., Marslen-Wilson, W. & Petkov, C. I. (2011). Behavioural and functional imaging analysis of "artificial-grammar" sequence learning in Rhesus macaques. *Society for Neuroscience,* 2011, Washington, DC, USA.

Wilson, B. & Petkov, C. I. (2010). Functional imaging of brain regions sensitive to communication sounds in primates. *Interspeech 2010*, Makuhari, Japan.

Kikuchi, Y., Barrett, J., Attaheri, A., Milne, A., **Wilson, B.** & Petkov, C. I. (2012). Neuroimaging and neurophysiology of Artificial Grammar Learning in the Primate Brain: Relationship between fMRI-BOLD and Neuronal Activity. *The Tucker-Davis Symposium on Advances and Perspectives in Auditory Neurophysiology (APAN)*, New Orleans, LA, USA

Bibliography

Abe, K. & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat Neurosci, 14*(8), 1067-1074.

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S. & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res, 166*(3-4), 559-571.

Amunts, K. & Zilles, K. (2012). Architecture and organizational principles of Broca's region. *Trends in Cognitive Sciences, 16*(8), 418-426.

Anwander, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D. & Knosche, T. R. (2007). Connectivity-based parcellation of Broca's area. *Cerebral Cortex, 17*(4), 816-825.

Arnold, K. & Zuberbuhler, K. (2006). Language evolution: semantic combinations in primate calls. *Nature, 441*(7091), 303.

Arnold, K. & Zuberbuhler, K. (2008). Meaningful call combinations in a non-human primate. *Curr Biol, 18*(5), R202-203.

Aslin, R. N., Woodward, J. Z., LaMendola, N. P. & Bever, T. G. (1996). Models of word segmentation in maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 117-134). Hillside, MJ: Erlbaum.

Aslin, R. N., Saffran, J. & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321-324.

Bahlmann, J., Schubotz, R. I. & Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *Neuroimage, 42*(2), 525-534.

Bahlmann, J., Schubotz, R. I., Mueller, J. L., Koester, D. & Friederici, A. D. (2009). Neural circuits of hierarchical visuo-spatial sequence processing. *Brain Res, 1298*, 161-170.

Bartholomew, G. A. & Collias, N. E. (1962). The role of vocalization in the social behaviour of the northern elephant seal. *Anim Behav, 10*, 7-8.

Bartolo, M. J., Baumann, S., Reser, D., Hunter, D., Griffiths, T. D., Sun, L., Rosa, M. G. P. & Thiele, A. (2009). *Reversible inactivation of the macaque auditory cortex with botulinum* *toxin - An fMRI time course analysis.* Paper presented at the Society for Neuroscience, Chicago, Il.

Bates, D., Maechler, M. Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. http://CRAN.R-project.org/package=lme4

Baumann, S., Griffiths, T. D., Rees, A., Hunter, D., Sun, L. & Thiele, A. (2010). Characterisation of the BOLD response time course at different levels of the auditory pathway in non-human primates. *Neuroimage, 50*, 1099-1108.

Bechara, A., Damasio, H. & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex, 10*(3), 295-307.

Bechara, A. & Van der Linden, M. (2005). Decision-making and impulse control after frontal lobe injuries. *Current Opinion in Neurology, 18*(6), 734-739.

Beckers, G. J. L., Bolhuis, J. J., Okanoya, K. & Berwick, R. C. (2012). Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport, 23*(3), 139-145.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P. & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Belin, P. & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport, 14*(16), 2105-2109.

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D. & Grodzinsky, Y. (2003). The neural reality of syntactic transformations: Evidence from functional magnetic resonance imaging. *Psychological Science, 14*(5), 433-440.

Ben-Shachar, M., Palti, D. & Grodzinsky, Y. (2004). Neural correlates of syntactic movement: converging evidence from two fMRI experiments. *Neuroimage, 21*(4), 1320-1336.

Bennett, C. L., Davis, R. T. & Miller, J. M. (1983). Demonstration of presbycusis across repeated measures in a nonhuman primate species. *Behavioral Neuroscience, 97*, 602-607.

Bergman, T. J., Beehner, J. C., Cheney, D. L. & Seyfarth, R. M. (2003). Hierarchical classification by rank and kinship in baboons. *Science, 302*(5648), 1234-1236.

Berwick, R. C., Okanoya, K., Beckers, G. J. & Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends Cogn Sci, 15*(3), 113-121.

Berwick, R. C., Beckers, G. J., Okanoya, K. & Bolhuis, J. J. (2012). A Bird's Eye View of Human Language Evolution. *Front Evol Neurosci, 4*, 5.

Bickerton, D. (2009). *Adam's Tounge: How humans made language, how language made humans*. New York, NY: Hill and Wang.

Bickerton, D. & Szathmary, E. (2009). *Biological Foundations and Origin of Syntax*. Cambridge, MA: MIT Press.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M. & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci, 17*(1), 353-362.

Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cereb Cortex*.

Bolhuis, J. J., Okanoya, K. & Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nat Rev Neurosci, 11*(11), 747-759.

Bornkessel, I., Zysset, S., Friederici, A. D., von Cramon, D. Y. & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage, 26*(1), 221-233.

Boughman, J. W. (1998). Vocal learning by greater spear-nosed bats. *Proc Biol Sci, 265*(1392), 227-233.

Bozic, M., Tyler, L. K., Ives, D. T., Randall, B. & Marslen-Wilson, W. D. Bihemispheric foundations for human speech comprehension. *Proc Natl Acad Sci U S A. 107*, 17439-17444.

Bozic, M., Tyler, L. K., Su. L., Wingfield, C & Marslen-Wilson, W. D. Neurobiological systems for lexical representation and analysis in English, *Journal of Cognitive Neuroscience, 25*(10) 1678-1691.

Brauer, J., Anwander, A. & Friederici, A. D. (2011). Neuroanatomical prerequisites for language functions in the maturing brain. *Cereb Cortex, 21*(2), 459-466.

Broca, P. (1861a). Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bulletin de la Société d'Anatomie, 6*, 330-357.

Broca, P. (1861b). Nouvelle observation d'aphe´mie produite par une le´sion de la troisie`me circonvolution frontale. *Bulletins de la Societe d'anatomie, 2*(6), 398-407.

Broca, P. (1861c). Perte de la parole: ramollissement chronique et destruction partielle du lobe anterieur gauche du cerveau. *Bulletins de la Societe d'anthropologie, 1*(2), 235-238.

Broca, P. (1865). Sur le siege de la faculte du langage articule. *Bulletin de la Societe d'anthropologie, 6*, 337-393.

Brodmann, K. (1909). Beiträge zur histologischen Lokalisation der Grosshirnrinde. VI. Die Cortexgliederung des Menschen. *J Psychol Neurol, 10*, 231-246.

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex, 11*(12), 1110-1123.

Candiotti, A., Zuberbuhler, K. & Lemasson, A. (2012a). Context-related call combinations in female Diana monkeys. *Animal Cognition, 15*(3), 327-339.

Candiotti, A., Zuberbuhler, K. & Lemasson, A. (2012b). Convergence and divergence in Diana monkey vocalizations. *Biol Lett, 8*(3), 382-385.

Caplan, D. & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*(1), 77-+.

Catani, M., Jones, D. K. & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann Neurol, 57*(1), 8-16.

Catani, M. & Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex, 44*(8), 1105-1132.

Catchpole, C. K. & Slater, P. J. B. (2008). *Bird Song*. Cambridge, UK: Cambridge University Press.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

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

Chomsky, N. (1972). *Language and Mind*: Harcourt, Brace and World.

Christiansen, M. H. & Kirby, S. (2003a). Language evolution: consensus and controversies. *Trends in Cognitive Sciences, 7*(7), 300-307.

Christiansen, M. H. & Kirby, S. (2003b). Language Evolution: The hardest problem in science? In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution*. Oxford: Oxford University Press.

Clarke, E., Reichard, U. H. & Zuberbuhler, K. (2006). The Syntax and Meaning of Wild Gibbon Songs. *PLoS ONE, 1*(1).

Clay, Z. & Zuberbuhler, K. (2009). Food-Associated Calling Sequences in Bonobos, Pan paniscus. *Folia Primatologica, 80*(2), 114-114.

Clay, Z. & Zuberbuhler, K. (2011). Bonobos Extract Meaning from Call Sequences. *PLoS ONE, 6*(4).

Clay, Z., Smith, C. L. & Blumstein, D. T. (2012). Food-associated vocalizations in mammals and birds: what do these calls really mean? *Animal Behaviour, 83*(2), 323-330.

Cole, R. A., Jakimik, J. & Cooper, W. E. (1980). Segmenting Speech into Words. Journal of the Acoustical Society of America, 67(4), 1323-1332.

Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W. J. & Shankweiler, D. (2004). Sentence complexity and input modality effects in sentence comprehension: an fMRI study. *Neuroimage, 22*(1), 11-21.

Cooke, A., Zurif, E. B., DeVita, C., Alsop, D., Koenig, P., Detre, J., Gee, J., Pinango, M., Balogh, J. & Grossman, M. (2002). Neural basis for sentence comprehension: Grammatical and short-term memory components. *Human Brain Mapping, 15*(2), 80-94.

Corballis, M. C. (2003). From hand to mouth: the gestural origins of language. In M. B. Christiansen & B. Kirby (Eds.), *Language Evolution*. Oxford: Oxford University Press.

Corballis, M. C. (2007). On phrase structure and brain responses: A comment on Bahlmann, Gunter, and Friederici (2006). *Journal of Cognitive Neuroscience, 19*(10), 1581-1583.

Crockford, C. & Boesch, C. (2005). Call combinations in wild chimpanzees. *Behaviour, 142*, 397-421.

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133–142.

Damasio, A. R. & Geschwind, N. (1984). The neural basis of language. *Annu Rev Neurosci, 7*, 127-147.

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., Barth, A. C. R., Maiworm, S., Knecht, S., Zwitserlood, P. & Floel, A. (2010). Electrical Stimulation of Broca's Area Enhances Implicit Learning of an Artificial Grammar. *Journal of Cognitive Neuroscience, 22*(11), 2427-2436.

Deacon, T. W. (1997). *The symbolic species*. New York: Penguin.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A. & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage, 24*(1), 21-33.

Dejerine, J. J. (1895). *Anatomie des Centres Nerveux*. Paris: Rueff et Cie.

Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S. & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage, 31*(3), 968-980.

DeVore, I. (1963). A comparison of the ecology and behaviour of monkeys and apes. In S. L. Washburn (Ed.), *Classification and human evolution*. Chicago, USA: Aldine.

Dittmar, M., Abbot-Smith, K., Lieven, E. & Tomasello, M. (2008). Young German children's early syntactic competence: a preferential looking study. *Developmental Science, 11*(4), 575-582.

Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T. & Cabanis, E. A. (2007). Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. *Brain, 130*(Pt 5), 1432-1441.

Dubois, J., Hertz-Pannier, L., Dehaene-Lambertz, G., Cointepas, Y. & Le Bihan, D. (2006). Assessment of the early organization and maturation of infants' cerebral white matter fiber bundles: A feasibility study using quantitative diffusion tensor imaging and tractography. *Neuroimage, 30*(4), 1121-1132.

Dunbar, R. B. (1998). The social brain hypothesis. *Evolutionary Anthropology, 6*, 178-190.

Eggert, G. H. (Ed.). (1977). *Wernicke's works on aphasia: a sourcebook and review*. The Hague: Mouton, 1977.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K. & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage, 25*(4), 1325-1335.

Elowson, A. E. & Snowdon, C. T. (1994). Pygmy marmosets,*Cebuella pygmaea*, modify vocal structure in response to changed social environment. *Animal Behaviour, 47*(6), 1267.

Ey, E., Rahn, C., Hammerschmidt, K. & Fischer, J. (2009). Wild Female Olive Baboons Adapt their Grunt Vocalizations to Environmental Conditions. *Ethology, 115*(5), 493-503.

Fedorenko, E., Gibson, E. & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language, 54*(4), 541-553.

Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E., Wada, K., Mouritsen, H. & Jarvis, E. D. (2008). Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS ONE, 3*(3), e1768.

Fiebach, C. J., Vos, S. H. & Friederici, A. D. (2004). Neural correlates of syntactic ambiguity in sentence comprehension for low and high span readers. *Journal of Cognitive Neuroscience, 16*(9), 1562-1575.

Fiebach, C. J., Schlesewsky, M., Lohmann, G., von Cramon, D. Y. & Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. *Human Brain Mapping, 24*(2), 79-91.

Field, A. P. (2009). Discovering statistics using SPSS. London, England : SAGE.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends Cogn Sci, 4*(7), 258-267.

Fitch, W. T. & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*(5656), 377-380.

Fitch, W. T., Hauser, M. D. & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition, 97*(2), 179-210.

Fitch, W. T. & Fritz, J. B. (2006). Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *J Acoust Soc Am, 120*(4), 2132-2141.

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge, UK: Cambridge University Press.

Fitch, W. T. & Friederici, A. (2012). Artificial grammar learning meets formal language theory: An overview. *Philos Trans R Soc Lond B Biol Sci, 367*, 1933-1955.

Folia, V., Udden, J., De Vries, M., Forkstam, C. & Petersson, K. M. (2010). Artificial language learning in adults and children. *Language Learning, 60*, 188-220.

Folia, V., Forkstam, C., Ingvar, M., Hagoort, P. & Petersson, K. M. (2011). Implicit Artificial Syntax Processing: Genes, Preference, and Bounded Recursion. *Biolinguistics, 5*, 105-132.

Fooden, J. (2000). Systematic review of the rhesus macaque, *Macaca mulatta* (Summermann, 1780). *Fieldiana Zoology, 96*, 1-180.

Forkstam, C., Hagoort, P., Fernandez, G., Ingvar, M. & Petersson, K. M. (2006). Neural correlates of artificial syntactic structure classification. *Neuroimage, 32*(2), 956-967.

Friederici, A. D., Meyer, M. & von Cramon, D. Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information (vol 74, pg 289, 2000). *Brain and Language, 75*(3), 465-477.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn Sci, 6*(2), 78-84.

Friederici, A. D., Steinhauer, K. & Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proc Natl Acad Sci U S A, 99*(1), 529-534.

Friederici, A. D., Ruschemeyer, S. A., Hahne, A. & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex, 13*(2), 170-177.

Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends Cogn Sci, 8*(6), 245-247.

Friederici, A. D. & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language, 89*(2), 267-276.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I. & Anwander, A. (2006a). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc Natl Acad Sci U S A, 103*(7), 2458-2463.

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D. & von Cramon, D. Y. (2006b). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cereb Cortex, 16*(12), 1709-1717.

Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends Cogn Sci, 13*(4), 175-181.

Friederici, A. D., Makuuchi, M. & Bahlmann, J. (2009). The role of the posterior superior temporal cortex in sentence comprehension. *Neuroreport, 20*(6), 563-568.

Friederici, A. D., Kotz, S. A., Scott, S. K. & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Hum Brain Mapp, 31*(3), 448-457.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol Rev, 91*(4), 1357-1392.

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.

Gentner, T. Q. & Hulse, S. H. (1998). Perceptual mechanisms for individual vocal recognition in European starlings, Sturnus vulgaris. *Anim Behav, 56*(3), 579-594.

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

Ghazanfar, A. A. & Hauser, M. D. (2001). The auditory behaviour of primates: a neuroethological perspective. *Curr Opin Neurobiol, 11*(6), 712-720.

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D. & Logothetis, N. K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr Biol, 17*(5), 425-430.

Gibson, E. & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences, 2*(7), 262-268.

Gifford, G. W., 3rd, MacLean, K. A., Hauser, M. D. & Cohen, Y. E. (2005). The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. *J Cogn Neurosci, 17*(9), 1471-1482.

Gil-da-Costa, R., Martin, A., Lopes, M. A., Munoz, M., Fritz, J. B. & Braun, A. R. (2006). Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. *Nat Neurosci, 9*(8), 1064-1070.

Gil, D. & Gahr, M. (2002). The honesty of bird song: multiple constraints for multiple traits. *Trends in Ecology & Evolution, 17*(3), 133-141.

Gomez, R. L. & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109-135.

Gomez, R. L. & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*(5), 178-186.

Gould, S. J. & Lewontin, R. C. (1979). Spandrels of San-Marco and the Panglossian Paradigm - a Critique of the Adaptationist Program. *Proceedings of the Royal Society Series B-Biological Sciences, 205*(1161), 581-598.

Grodzinsky, Y. & Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. *Curr Opin Neurobiol, 16*(2), 240-246.

Grodzinsky, Y. & Santi, A. (2008). The battle for Broca's region. *Trends in Cognitive Sciences, 12*(12), 474-480.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences, 9*(9), 416-423.

Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A. & Summerfield, A. Q. (2002). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 12*(2), 140-149.

Harcourt, A. H. & Stewart, K. J. (1996). Function and meaning of wild gorilla 'close' calls. *Behaviour, 133*, 827-845.

Hauser, M. D. (1993). Rhesus monkey copulation calls: honest signals for female choice? *Proc Biol Sci, 254*(1340), 93-96.

Hauser, M. D. & Marler, P. (1993a). Food-associated calls in rhesus macaques (*Macaca mulatta*): II. Costs and benefits of call production and suppression. *Behavioral Ecology,, 4*(3), 206-212.

Hauser, M. D. & Marler, P. (1993b). Food-associated calls in rhesus macaques (*Macaca mulatta*): I. Socioecological factors. *Behavioral Ecology, 4*(3), 194-205.

Hauser, M. D., Newport, E. L. & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition, 78*(3), B53-64.

Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Hauser, M. D., Barner, D. & O'Donnell, T. (2007). Evolutionary linguistics: A new look at the old landscape. *Language and Learning and development, 3*(2), 101-132.

Hauser, M. D. & Glynn, D. (2009). Can Free-Ranging Rhesus Monkeys (Macaca mulatta) Extract Artificially Created Rules Comprised of Natural Vocalizations? *Journal of Comparative Psychology, 123*(2), 161-167.

Hickok, G. & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1-2), 67-99.

Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nat Rev Neurosci, 8*(5), 393-402.

Honda, E. & Okanoya, K. (1999). Acoustical and syntactical comparisons between songs of the white-backed Munia (Lonchura striata) and its domesticated strain, the Bengalese finch (Lonchura striata var. domestica). *Zoological Science, 16*, 319-326.

Houston, S. D. (2008). *The first writing: Script invention as history and process.* Cambridge: Cambridge University Press.

Hubrecht, R. C. (1984). Field observations on group size and composition of the common marmoset (Callithrix jacchus jacchus), at Tapacura, Brazil. *Primates, 25*(1), 13-21.

Hudson Kam, C. L. & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59*(1), 30-66.

Humboldt, W. von. (1836). Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwickelung des Menschengeschlechts. Berlin: F. Dummler.

Humphries, C., Love, T., Swinney, D. & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Hum Brain Mapp, 26*(2), 128-138.

Humphries, C., Binder, J. R., Medler, D. A. & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J Cogn Neurosci, 18*(4), 665-679.

Hurford, J. R. (2012). *The Origins of Grammar: Language in the Light of Evolution*: Oxford University Press.

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences, 3*(7), 272-279.

Jackendoff, R. (2002). *Foundations of Language.* Oxford: Oxford University Press.

Jackendoff, R. & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language - (Reply to Fitch, Hauser, and Chomsky). *Cognition, 97*(2), 211-225.

Jackson, L. L., Heffner, R. S. & Heffner, H. E. (1999). Free-field audiogram of the Japanese macaque (Macaca fuscata). *Journal of the Acoustical Society of America, 106*(5), 3017-3023.

Jaeger, G. & Rogers, J. (2012). Formal language theory: Refining the Chomsky hierarchy. *Philos Trans R Soc Lond B Biol Sci, 367*, 1956-1970.

Janik, V. M. & Slater, P. J. (1997). Vocal learning in mammals. *Advances in the Study of Behavior, 26*, 59-99.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Ann N Y Acad Sci, 1016*, 749-777.

Just, M. A. & Carpenter, P. A. (1992). A Capacity Theory of Comprehension - Individual-Differences in Working Memory. *Psychological Review, 99*(1), 122-149.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F. & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science, 274*(5284), 114-116.

Karnath, H. O., Ferber, S. & Himmelbach, M. (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature, 411*(6840), 950-953.

Kinno, R., Kawamura, M., Shioda, S. & Sakai, K. L. (2008). Neural correlates of noncanonical syntactic processing revealed by a picture-sentence matching task. *Human Brain Mapping, 29*(9), 1015-1027.

Koda, H., Shimooka, Y. & Sugiura, H. (2008). Effects of Caller Activity and Habitat Visibility on Contact Call Rate of Wild Japanese Macaques (Macaca fuscata). *American Journal of Primatology, 70*(11), 1055-1063.

Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., Patterson, R. D., Howard, M. A., Friston, K. J., Griffiths, T. D. (2011). Predictive coding and pitch processing in the auditory cortex. *Journal of Cognitive Neuroscience, 23*(10), 3084-3094

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Lemasson, A. & Hausberger, M. (2004). Patterns of vocal sharing and social dynamics in a captive group of Campbell's monkeys (Cercopithecus campbelli campbelli). *Journal of Comparative Psychology, 118*(3), 347-359.

Lemasson, A., Richard, J.-P. & Hausberger, M. (2004). A new methodological approach to context analysis of call production. *Bioacoustics: The International Journal of Animal Sound and its Recording, 14*(2), 111-125.

Lemasson, A. & Hausberger, M. (2011). Acoustic variability and social significance of calls in female Campbell's monkeys (*Cercopithecus campbelli campbelli*). *Journal of Acoustical Society of America, 129*(8), 3341-3352.

Lewis, R. L., Vasishth, S. & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*(10), 447-454.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T. & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cereb Cortex, 15*(10), 1621-1631.

Logothetis, N. K., Guggenberger, H., Peled, S. & Pauls, J. (1999). Functional imaging of the monkey brain. *Nat Neurosci, 2*(6), 555-562.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature, 453*(7197), 869-878.

Lonsbury-Martin, B. L. & Martin, G. K. (1981). Effects of moderately intense sound on auditory sensitivity in rhesus monkeys: Behavioral and neural observations. *Journal of Neurophysiology, 46*, 563-586.

MacDougall-Shackleton, S. A. (1997). Sexual selection and the evolution of song repertoires. *Current Ornithology, 14*, 81-125.

Makuuchi, M., Bahlmann, J., Anwander, A. & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proc Natl Acad Sci U S A, 106*(20), 8362-8367.

Marcus, G. F., Vijayan, S., Bandi Rao, S. & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77-80.

Marler, P. & Tenaza, R. (1977). Communication in apes with special reference to vocalizations. In T. A. Sebeok (Ed.), *How Animals Communicate* (pp. 965-1033). Bloomington: Indiana University Press.

Marler, P. (2000). In N. L. Wallin, B. Merker & S. Brown (Eds.), *The Origins of Music* (pp. 31-50): Massachusetts Institute of Technology.

Marslen-Wilson, W., Bozic, M. & Tyler, L. K. in *Cognitive Neuroscience (5th Ed)*. Eds M. Gazzaniga, G. Davies, R. B. Ivry, & G.R. Mangun. MIT Press, in press.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L. & Mehler, J. (1993). The Cortical Representation of Speech. *Journal of Cognitive Neuroscience, 5*(4), 467-479.

McComb, K. E. (1991). Female Choice for High Roaring Rates in Red Deer, Cervus-Elaphus. *Animal Behaviour, 41*, 79-88.

McLaren, D. G., Kosmatka, K. J., Oakes, T. R., Kroenke, C. D., Kohama, S. G., Matochik, J. A., Ingram, D. K. & Johnson, S. C. (2009). A population-average MRI-based atlas collection of the rhesus macaque. *Neuroimage, 45*(1), 52-59.

Miles, R. C. & Meyer, D. C. (1956). Learning sets in marmosets. *Journal of Comparative and Physiological Psychology, 49*, 212-222.

Miles, R. C. (1957). Delayed-response learning in the marmoet and the macaque. *Journal of Comparative Physiological Psychology, 50*(4), 352-355.

Miller, G. F. (2001). The mating mind: How sexual choice shaped the evolution of human nature. *Psycoloquy, 12*(8), 1-15.

Moore, B. R. (2004). The evolution of learning. *Biol Rev Camb Philos Soc, 79*(2), 301-335.

Murphy, R. A., Mondragon, E. & Murphy, V. A. (2008). Rule learning by rats. *Science, 319*(5871), 1849-1851.

Newmeyer, F. (2003). What Can the Field of Linguistics Tell Us About the Origins of Language?" In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution*, Oxford: Oxford University Press.

Newport, E. L. & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn Psychol, 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. *Cogn Psychol, 49*(2), 85-117.

Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N. & Jenner, K. C. (2000). Cultural revolution in whale songs. *Nature, 408*(6812), 537.

Obleser, J., Scott, S. K. & Eulitz, C. (2006). Now you hear it, now you don't: transient traces of consonants and their nonspeech analogues in the human brain. *Cereb Cortex, 16*(8), 1069-1076.

Obleser, J., Wise, R. J., Alex Dresner, M. & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci, 27*(9), 2283-2289.

Obleser, J. & Kotz, S. A. (2010). Expectancy Constraints in Degraded Speech Modulate the Language Comprehension Network. *Cerebral Cortex, 20*(3), 633-640.

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H. & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A, 89*(13), 5951-5955.

Okano, H., Hikishima, K., Iriki, A. & Sasaki, E. (2012). The common marmoset as a novel animal model system for biomedical and neuroscience research applications. *Semin Fetal Neonatal Med, 17*(6), 336-340.

Okanoya, K. (2004a). Song syntax in bengalese finches: Proximate and ultimate analyses. *Advances in the Study of Behavior, Vol 34, 34*, 297-346.

Okanoya, K. (2004b). The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. *Ann N Y Acad Sci, 1016*, 724-735.

Opitz, B. & Friederici, A. D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *Neuroimage, 19*(4), 1730-1737.

Opitz, B. & Friederici, A. D. (2004). Brain correlates of language learning: the neuronal dissociation of rule-based versus similarity-based learning. *J Neurosci, 24*(39), 8436-8440.

Ouattara, K., Lemasson, A. & Zuberbuhler, K. (2009a). Campbell's monkeys use affixation to alter call meaning. *PLoS ONE, 4*(11), e7808.

Ouattara, K., Lemasson, A. & Zuberbuhler, K. (2009b). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proc Natl Acad Sci U S A, 106*(51), 22026-22031.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S. & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36*(4), 767-776.

Pena, M., Bonatti, L. L., Nespor, M. & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*(5593), 604-607.

Pepperberg, I. M. (2010). Vocal learning in Grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain Lang, 115*(1), 81-91.

Perani, D., Saccuman, M. C., Scifo, P., Awander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G. & Friederici, A. D. (2011). Neural language networks at birth. *Proc Natl Acad Sci U S A, 108*(38), 16056-16061.

Perrodin, C., Kayser, C., Logothetis, N. K. & Petkov, C. I. (2011). Voice cells in the primate temporal lobe. *Curr Biol, 21*(16), 1408-1415.

Perruchet, P. & Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review, 12*(2), 307-313.

Petersson, K. M., Forkstam, C. & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science, 28*(3), 383-407.

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.

Petersson, K. M. & Hagoort, P. (2012). The neurobiology of syntax: beyond string sets. *Philosophical Transactions of the Royal Society B-Biological Sciences, 367*(1598), 1971-1983.

Petkov, C. I., Kayser, C., Augath, M. & Logothetis, N. K. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol, 4*(7), e215.

Petkov, C. I., Kayser, C., Ghazanfar, A., Patterson, R. D. & Logothetis, N. K. (2008a). *Functional imaging of sensitivity to components of the voice in monkey auditory cortex.* Paper presented at the Society for Neuroscience, Washington D.C.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M. & Logothetis, N. K. (2008b). A voice region in the monkey brain. *Nat Neurosci, 11*(3), 367-374.

Petkov, C. I., Kikuchi, Y., Augath, M., Mishkin, M., Rauschecker, J. & Logothetis, N. K. (2008c). *Voice region connectivity in the monkey assessed with microstimulation and functional imaging.* Paper presented at the Satellite meeting to Society for Neuroscience: Advances and Perspectives in Auditory Neurophysiology (APAN), APAN satellite meeting to Society for Neuroscience.

Petkov, C. I., Kayser, C., Augath, M. & Logothetis, N. K. (2009). Optimizing the imaging of the monkey auditory cortex: sparse vs. continuous fMRI. *Magn Reson Imaging, 27*(8), 1065-1073.

Petkov, C. I. & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience, 4*, 12.

Petkov, C. I. & Wilson, B. (2012). On the pursuit of the brain network for proto-syntactic learning in nonhuman primates: Conceptual issues and neurobiological hypotheses. *Philos Trans R Soc Lond B Biol Sci, 367*, 2077-2088.

Pfingst, B. E., Hienz, R. & Miller, J. (1975). Reaction-Time Procedure for Measurement of Hearing .2. Threshold Functions. *Journal of the Acoustical Society of America, 57*(2), 431-436.

Pfingst, B. E., Laycock, J., Flammino, F., Lonsburymartin, B. & Martin, G. (1978). Pure-Tone Thresholds for Rhesus-Monkey. *Hearing Research, 1*(1), 43-47.

Pinker, S. & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13*, 707-784.

Pinker, S. (1994). *The Language Instinct.* New York: William Morrow and Company, Inc.

Pinker, S. & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition, 95*(2), 201-236.

Poirier, C., Boumans, T., Verhoye, M., Balthazart, J. & Van der Linden, A. (2009). Own-song recognition in the songbird auditory pathway: selectivity and lateralization. *J Neurosci, 29*(7), 2252-2258.

Pola, Y. V. & Snowdon, C. T. (1975). Vocalizations of Pygmy Marmosets (Cebuella-Pygmaea). *Animal Behaviour, 23*(Nov), 826-842.

Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E., Herscovitch, P. & Mishkin, M. (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature, 427*(6973), 448-451.

Pulvermuller, F. & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience, 11*(5), 351-360.

Rauschecker, J. P. & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci, 12*(6), 718-724.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour, 6*(6), 855-863.

Rendall, D., Owren, M. J. & Rodman, P. S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (Macaca mulatta) vocalizations. *J Acoust Soc Am, 103*(1), 602-614.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X. & Behrens, T. E. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat Neurosci, 11*(4), 426-428.

Rimol, L. M., Specht, K., Weis, S., Savoy, R. & Hugdahl, K. (2005). Processing of sub-syllabic speech units in the posterior temporal lobe: an fMRI study. *Neuroimage, 26*(4), 1059-1067.

Rizzolatti, G. & Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci, 21*(5), 188-194.

Rodd, J. M., Davis, M. H. & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex, 15*(8), 1261-1269.

Rogalsky, C., Matchin, W. & Hickok, G. (2008). Broca's area, sentence comprehension, and working memory: an fMRI study. *Front Hum Neurosci, 2.*

Rogalsky, C. & Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cereb Cortex, 19*(4), 786-796.

Rogalsky, C. & Hickok, G. (2011). The Role of Broca's Area in Sentence Comprehension. *Journal of Cognitive Neuroscience, 23*(7), 1664-1680.

Saffran, J., Newport, E. L. & Aslin, R. N. (1996a). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.

Saffran, J., Newport, E. L., Aslin, R. N., Tunick, R. A. & Barrueco, S. (1997). Incidental language learing: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 101-105.

Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F. & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition, 107*(2), 479-500.

Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996b). 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.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*(1), 172-196.

Saleem, K. S. & Logothetis, N. K. (2007). *A Combined MRI and Histology: Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates*. London: Academic Press.

Sanides, F. (1962). *Die Architektonik des menschlichen Stirnhirns*. Berlin: Springer.

Santi, A. & Grodzinsky, Y. (2007). Working memory and syntax interact in Broca's area. *Neuroimage, 37*(1), 8-17.

Santi, A. & Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *Neuroimage, 51*(4), 1285-1293.

Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M. S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J. & Weiller, C. (2008). Ventral and dorsal pathways for language. *Proc Natl Acad Sci U S A, 105*(46), 18035-18040.

Saur, D., Schelter, B., Schnell, S., Kratochvil, D., Kupper, H., Kellmeyer, P., Kummerer, D., Kloppel, S., Glauche, V., Lange, R., Mader, W., Feess, D., Timmer, J. &

Weiller, C. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *Neuroimage, 49*(4), 3187-3197.

Schel, A. M., Candiotti, A. & Zuberbuhler, K. (2010). Predator-deterring alarm call sequences in Guereza colobus monkeys are meaningful to conspecifics. *Animal Behaviour, 80*(5), 799-808.

Schmitt, D. (2010). Primate locomotor evolution: Biomechanical studies of primate locomotion and their implications for understanding primate neuroethology. In A. A. Ghazanfar & M. L. Platt (Eds.), *Primate Neuroethology* (pp. 10-30): Oxford University Press.

Searcy, W. A. & Anderson, M. (1986). Secual selection and the evolution of song. *Annual review of Ecology and Systematics, 17*, 507-533.

Seiden, H. R. (1958). *Auditory acuity of the marmoset monkey (Hapale jacchus).* Unpublished Doctoral Thesis, Princeton University, Princeton University

Seiler, M., Schwitzer, C., Gamba, M. & Holderied, M. W. (2013). Interspecific Semantic Alarm Call Recognition in the Solitary Sahamalaza Sportive Lemur, *Lepilemur sahamalazensis*. *PLoS ONE, 8*(6).

Seyfarth, D. & Cheney, D. (1984). The acoustic features of vervet monkey grunts. *Journal of Acoustical Society of America, 75*(5), 1623-1628.

Seyfarth, R. M., Cheney, D. L. & Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science, 210*(4471), 801-803.

Seyfarth, R. M. & Cheney, D. L. (1999). Production, usage, and response in nonhuman primate vocal development. In M. D. Hauser & M. Konishi (Eds.), *The Design of Animal Communication* (pp. 391-417). Cambridge, Massachusetts: MIT Press.

Slocombe, K. E. & Zuberbuhler, K. (2005). Functionally referential communication in a chimpanzee. *Curr Biol, 15*(19), 1779-1784.

Slocombe, K. E., Kaller, T., Turman, L., Townsend, S. W., Papworth, S., Squibbs, P. & Zuberbuhler, K. (2010). Production of food-associated calls in wild male chimpanzees is dependent on the composition of the audience. *Behavioral Ecology and Sociobiology, 64*(12), 1959-1966.

Smith, K. & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. . *Cognition, 116*(3), 444-449.

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M. & Hagoort, P. (2009). Retrieval and Unification of Syntactic Structure in Sentence Comprehension: an fMRI Study Using Word-Category Ambiguity. *Cerebral Cortex, 19*(7), 1493-1503.

Snijders, T. M., Petersson, K. M. & Hagoort, P. (2010). Effective connectivity of cortical and subcortical regions during unification of sentence structure. *Neuroimage, 52*(4), 1633-1644.

Snowdon, C. T., Elowson, A. E. & Roush, R. S. (1997). Social influences on vocal development in New World primates. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development*. Cambridge: Cambridge University Press.

Snowdon, C. T. & Elowson, A. E. (1999). Pygmy marmosets modify call structure when paired. *Ethology, 105*(10), 893-908.

Snowdon, C. T. (2009). Plasticity of communication in nonhuman primates. In M. Naguib & V. M. Janik (Eds.), *Advances in the Study of Behavior* (Vol. 40, pp. 239-276). Burlington: Academic Press.

Steiper, M. E. & Young, N. M. (2006). Primate molecular divergence dates. *Molecular Phylogenetics and Evolution, 41*(2), 384-394.

Stobbe, N., Westphal-Fitch, G., Aust, U. & Fitch, W. T. (2012). Visual artificial grammar learning: comparative research on humans, kea (Nestor notabilis) and pigeons (Columba livia). *Philos Trans R Soc Lond B Biol Sci, 367*(1598), 1995-2006.

Stromswold, K., Caplan, D., Alpert, N. & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language, 52*(3), 452-473.

Struhsaker, T. T. (1967). Auditory communication among vervet monkeys (*Cercopithecus aethiops*). In S. A. Altman (Ed.), *Social communication among primates* (pp. 281-324). Chicago: University of Chicago Press.

Suge, R. & Okanoya, K. (2010). Perceptual chunking in the self-produced songs of Bengalese finches (Lonchura striata var. domestica). *Animal Cognition, 13*(3), 515-523.

Sun, J. X., Helgason, A., Masson, G., Ebenesersdottir, S. S., Li, H., Mallick, S., Gnerre, S., Patterson, N., Kong, A., Reich, D. & Stefansson, K. (2012). A direct characterization of human mutation based on microsatellites. *Nat Genet, 44*(10), 1161-1165.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A. & Hopkins, W. D. (2008). Communicative signaling activates 'Broca's' homolog in chimpanzees. *Curr Biol, 18*(5), 343-348.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A. & Hopkins, W. D. (2009). Visualizing vocal perception in the chimpanzee brain. *Cereb Cortex, 19*(5), 1151-1157.

Tallerman, M. (2011). What is syntax? In M. Tallerman & K. R. Gibson (Eds.), *The Oxford Handbook of Language Evolution* (pp. 442-455): Oxford University Press.

Teal, T. & Taylor, C. E. (2000). Effects of compression on language evolution *Artificial Life* (Vol. 6 (2), pp. 129-143): MIT Press.

ten Cate, C. & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philos Trans R Soc Lond B Biol Sci, 367*, 1984-1994.

Tian, B., Reser, D., Durham, A., Kustov, A. & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science, 292*(5515), 290-293.

Tyler, L. K. & Marslen-Wilson, W. (2008). Fronto-temporal brain systems supporting spoken language comprehension. *Philos Trans R Soc Lond B Biol Sci, 363*(1493), 1037-1054.

Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., Papoutsi, M. & Stamatakis, E. A. (2011). Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain, 134*(Pt 2), 415-431.

Udden, J., Folia, V., Forkstam, C., Ingvar, M., Fernandez, G., Overeem, S., van Elswijk, G., Hagoort, P. & Petersson, K. M. (2008). The inferior frontal cortex in artificial syntax processing: an rTMS study. *Brain Res, 1224*, 69-78.

Udden, J., Ingvar, M., Hagoort, P. & Petersson, K. M. (2011). Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: Investigating the push-down stack model. *Cognitive Science*.

Uster, D. & Zuberbuhler, K. (2001). The functional significance of Diana monkey 'clear' calls. *Behaviour, 138*, 741-756.

van Heijningen, C. 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 U S A, 106*(48), 20538-20543.

Van Meir, V., Boumans, T., De Groof, G., Van Audekerke, J., Smolders, A., Scheunders, P., Sijbers, J., Verhoye, M., Balthazart, J. & Van der Linden, A. (2005). Spatiotemporal properties of the BOLD response in the songbirds' auditory circuit during a variety of listening tasks. *Neuroimage, 25*(4), 1242-1255.

Vandenberghe, R., Nobre, A. C. & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience, 14*(4), 550-560.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B. & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage, 30*(4), 1414-1432.

von Kriegstein, K., Eger, E., Kleinschmidt, A. & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res Cogn Brain Res, 17*(1), 48-55.

Wager, T. D. & Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-anaylsis. *Cogn Affect Behav Neurosci, 3*(4), 255-274.

Waters, G. S. & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review, 103*(4), 761-772.

Weiss, D. J., Garibaldi, B. T. & Hauser, M. D. (2001). The production and perception of long calls by cotton-top tamarins (*Saguinus oedipus*): Acoustic analyses and playback experiments. *Journal of Comparative Psychology, 115*(3), 258-271.

Wernicke, C. (1874). Der aphasische Symptomenkomplex. In G. H. Eggert (Ed.), *Wernicke's works on aphasia : a sourcebook and review* (pp. 91-145). The Hague: Mouton, 1977.

Wilson, B. & Petkov, C. (2011). Communication and the primate brain: Insights from neuroimaging studies in humans, chimpanzees and macaques. *Human Biology, 83*(2), 14.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. http://arxiv.org/pdf/1308.5499.pdf

Wright, P., Stamatakis, E. A. & Tyler, L. K. Differentiating Hemispheric Contributions to Syntax and Semantics in Patients with Left-Hemisphere Lesions. *Journal of Neuroscience, 32*, 8149-8157

Zahavi, A. (1975). Mate Selection - Selection for a Handicap. *Journal of Theoretical Biology, 53*(1), 205-214.

Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*(5058), 846-849.

Zatorre, R. J., Bouffard, M., Ahad, P. & Belin, P. (2002). Where is 'where' in the human auditory cortex? *Nat Neurosci, 5*(9), 905-909.

Zoloth, S. R., Petersen, M. R., Beecher, M. D., Green, S., Marler, P., Moody, D. B. & Stebbins, W. (1979). Species-specific perceptual processing of vocal sounds by monkeys. *Science, 204*(4395), 870-873.

Zuberbuhler, K. (2000a). Referential labelling in Diana monkeys. *Anim Behav, 59*(5), 917-927.

Zuberbuhler, K. (2000b). Causal cognition in a non-human primate: field playback experiments with Diana monkeys. *Cognition, 76*(3), 195-207.

Zuberbuhler, K. (2000c). Interspecies semantic communication in two forest primates. *Proc Biol Sci, 267*(1444), 713-718.

Zuberbuhler, K. (2000d). Causal knowledge of predators' behaviour in wild Diana monkeys. *Anim Behav, 59*(1), 209-220.

Zuberbuhler, K. (2002). A syntactic rule in forest monkey communication. *Anim Behav, 63*, 293-299.

Zuberbuhler, K. & Jenny, D. (2002). Leopard predation and primate evolution. *J Hum Evol, 43*(6), 873-886.