Complexity of Primate Cognition through Audio-Visual Sensory Convergence

Zhipeng Zhang

Submitted for the degree of Doctor of Philosophy

Biosciences Institute

Newcastle University

July 2024

General Abstract

Language allows humans to flexibly create complex mental constructions of meaningful information from any sensory input. Although nonhuman primate communication lacks the full combinatorial capacity of human language, their cognitive system is known to be sophisticated for integrating various forms of information from sensory inputs into meaningful representation. One of the objectives of this thesis is to determine whether macaque monkeys can combine sensory information from sounds to identify visual objects by two properties (colour and shape). We designed a novel behavioural touchscreen task that was implemented with rhesus monkeys (*Macaca mulatta*) in their home units to study whether the monkeys could learn the meaning of speech sounds and use these to identity objects by their colour and shape. The task was designed into 3 phases including, 1) colour or shape associative learning; 2) alternation between colour and shape learning; and 3) combinatorial learning of colour and shape integrations. The permutation test results used to identify above chance performance, revealed that the monkeys learnt to associate the nonsense words with the colour and shape properties of the visual stimuli although consistently performing the task at high levels was difficult for both animals. Furthermore, one of the monkeys was tested on its ability to integrate sound information about both colour and shape and performed very well on the final phase of testing. This monkey provided important insights into the complexity of primate cognitive abilities as a basis for better understanding language evolution. The results from this work, suggest that basic human combinatorial abilities may be shared with nonhuman primates, although more animals need to be tested. Additionally, nonhuman primate and human brains appear to share multisensory integration abilities. I also completed data collection for a human and monkey functional MRI study on how the brain processes the temporal structure of audio-visual sequences. This study is added as an appendix to this thesis and will serve as the basis for future analyses to understand how this neural system for audio-visual sensory convergence compares to humans.

COVID-19 Impact Statement

I started my PhD in September 2019. However, the lab work was suspended in March 2020 due to the pandemic, as access to the animal unit was limited to protect the animals from the high possibility of infection, especially the macaque monkeys that were involved in my PhD work. My work was not prioritised and only the essential work was permitted in the animal facility (Comparative Biology Centre). This caused a significant delay in my experiments for approximately 5 months, and I did not have a chance to get back to the lab to carry on my PhD work until August 2020.

It was also challenging to work remotely during the pandemic at the early stage of my PhD, while I was working on the experimental design needed to be tested locally to make sure everything worked. Moreover, it took more time than expected to get the animals back on training and testing after I went back to the lab, and their behaviours also varied in each individual animal. This also caused a significant delay of the experimental data collection afterwards and I did not complete the data collection until July 2023.

In addition, I am an international student and travelled back home to China when it became possible during the pandemic. This added unforeseen delay to my PhD work, as I had to go through the strict rules and quarantine in China and the UK (October 2021 – January 2022). I had hoped to work remotely, but there were issues with access to the fileserver and Linux machine while I was in China, although this was working well when I tested it before I left Newcastle. Therefore, I could not make any further progress on data analysis as expected during the time of my visit to China, and this further delayed my PhD work by at least 3 months.

Finally, I had COVID in October 2022, and the post-Covid symptoms lasted for 3 weeks, which also caused the delay in my PhD work.

Acknowledgements

I would like to thank many people who offered me the help and support during my PhD.

I would fist thank my supervisor Chris Petkov and my co-supervisor Yukiko Kikuchi for their unwavering guidance and support. I would also like to thank my co-supervisor Tim Griffiths for his help.

I would also like to thank Ben Wilson and Fenna Poletiek for the partnership in the touchscreen project for their invaluable insight and help. On top of that, I would like to thank Ross Muers and Alice Milne for their help and support in the comparative fMRI project.

Thanks must go to past and present lab members, Jen Nacef, Ben Slater, Cody McCants, David Howett, Holly Jenkins, Beshoy Agayby and Emma Woolgar, whom it has been a great pleasure to work with. Especially Ben and Jen who helped me a lot during the long journey.

I would also like to thank Fabien Balezeau and David Hunter who offered me guidance and support for the fMRI project. Additionally, I would like to thank Ian Milne who offered me help with improvements of the touchscreen apparatus and many people in the Comparative Biology Centre who have also offered me the help and support.

Finally, I must say a big thank you to my parents who have supported and helped me emotionally and practically through the long journey. I would also like to thank my grandparents for their understanding and support. I am grateful for having you in my life.

Statement of Credit

This thesis was written entirely by me, with input, guidance, some minor editing and comments from my supervisors Chris Petkov and Yukiko Kikuchi.

Behavioural experiments: I designed the task with Ben Wilson, Fenna Poletiek and Chris Petkov. I conducted all of the behavioural testing, with support with the analysis and experimental design and execution from Ben Wilson and Chris Petkov wherever needed. I collected the data and conducted all of the analysis, with some analytical support from Chris Petkov & Ben Wilson as needed.

fMRI experiments: The fMRI experiments were a part of a long running comparative human and monkey neuroimaging project in the laboratory, started by Alice Milne, who collected much of the human data. Then the project was pushed forward by Ross Muers who collected one of the monkey's datasets and completed the human data collection. Because this work on audio-visual integration was highly relevant to my thesis work, I completed the data collection for the second monkey to round out the data acquisition for the human and monkey fMRI work and to give me experience with fMRI. I conducted the pre-processing of these datasets. However, because the behavioural work was so demanding and took the priority of my time during my thesis, the analysis of these datasets could not be completed during the thesis work and will continue in the future with me involved. For this reason, I append the fMRI project design and basis as well as a summary of the data collected as an appendix to ensure continuity with this work.

Table of Contents

List of Figures

List of Tables

Abbreviations

AG	Artificial Grammar
AL	The Anterior Lateral Area
ANOVA	The Analysis of Variance
aSTG	Anterior Superior Temporal Gyrus
AV	Audio-Visual
A1	Primary Auditory Cortex
A2	The Second Auditory Cortex
A3	The Tertiary Auditory Cortex
CANTAB	Cambridge Neuropsychological Test Automated Battery
CBC	The Comparative Biology Centre
CI	Confidence Interval
FG	Fusiform Gyrus
fMRI	Functional Magnetic Resonance Imaging
FOP	Frontal Operculum
GLM	General Linear Model
IT	Inferior Temporal Cortex
ML	The Middle Lateral Area
MT	The Medial Temporal Area
NHPs	Nonhuman Primates
PIT	The Posterior Inferotemporal Area
PFC	Prefrontal Cortex
RDM	Representational Dissimilarity Matrix
RSA	Representational Similarity Analysis
SD	Standard Deviations
SSRT	The Stop Signal Reaction Time task
STG	Superior Temporal Gyrus

STS	Superior Temporal Sulcus
TP	The Transitional Probabilities
VLPFC	Ventrolateral Prefrontal Cortex
V1	Primary Visual Cortex
V2	Visual Area 2
V4	Visual Area 4
V5	Visual Area 5
V6	Visual Area 6
WCST	The Wisconsin Card Sorting Test

Chapter 1: General Introduction

1.1 Language, Semantics, and Syntax

Human language is unique in its full combinatorial capacity distinguishing humans from all other animals. Human language involves a complex interplay of semantics (the meaningful language symbols that we use) and syntax (the grammar of language). By integrating symbols into grammatically structured sentences, we can generate an infinite number of expressions, whereas most animals have more limited expression or lack the full complexity of human language (Noam Chomsky, 2014; Trask & Stockwell, 2007). Moreover, language is flexibly applied to many forms of sensory input. Specifically, language can be formed in spoken, written and sign languages, and it is structured in grammar and vocabulary, by which can be conveyed into meaning by human beings (Crystal & Crystal, 1981; Trask & Stockwell, 2007).

Syntax and semantics play important roles in language, by which these processes order meaningful words into grammatical sentences (Akmajian, Farmer, Bickmore, Demers, & Harnish, 2017). More specifically, semantics provides the meaning of the elements in a sequence. For example, in the sentence 'The talk at Newcastle was fascinating', each of the words carries meaning as does the structure of the sentence, which in this example indicates that the talk was fascinating rather than Newcastle, necessarily. When we talk about the interplay of semantics and syntax, this typically underlies the order of words in which a sentence is structured, and the rules of the syntax give the sentence greater complexity and meaning (Wilson et al., 2013). However, it remains unclear which other animals, including nonhuman primates (NHPs) share similarities with humans with regard to aspects of language processing.

With regard to semantics, the general principle is that many nonhuman animals are able to use and comprehend meaningful communications from other animals. For example, Cheney and Seyfarth showed the evidence of semantic communications between monkeys by conducting an experiment of their responses to different alarm calls. Their results indicated that monkeys behaved differently to leopard, eagle and snake alarms by running into trees, looking up and looking down which suggest that these varied alarm calls have different meanings to them, as such the semantic communications are present among the monkeys (R. M. Seyfarth, Cheney, & Marler, 1980). Similar findings also indicated that the monkeys could produce alarm calls to identify and refer to different predators such as eagles and leopards (Zuberbühler, 2000, 2001). Interestingly, even among different species of monkeys, Diana monkeys could interpret eagle-alarm calls produced by another species, putty-nosed monkeys as the presence of eagle (Eckardt

& Zuberbühler, 2004), suggesting that the vocal communication among monkeys might carry different meanings referring to a subject or an event, and that individuals within and outside of the species can learn to recognise their meaning to respond appropriately.

Not only are NHPs able to produce and comprehend vocalisations that serve as predator alarm calls, but they can also integrate vocalisations into meaningful sets of units with basic combinatorial operations (Arnold & Zuberbühler, 2006). This means that a basic facet of both semantic and syntactic capabilities is shared with NHPs, even if the full complexity of language syntax, often referred to as the full combinatorial or full generative capability of language, may be human unique. Similarly, songbirds can produce song sequences that vary in the complexity of the song structure, although even this complexity is not viewed to be as complex as that for human language which lacks the complexity of syntactic structures in language (Berwick, Okanoya, Beckers, & Bolhuis, 2011; Fitch & Jarvis, 2013). Other studies including vocal learning and discrimination and production of centre-embedded pattern have also demonstrated that cognitive processes such as perception, decision-making and production of language are also present in NHPs (Doupe & Kuhl, 1999; Jiang et al., 2018).

The human brain is not likely to be 'installed' with language at birth, although its neural system might be predisposed to it. Noam Chomsky has eloquently argued that language is innate and likely genetically predetermined as a neural system at birth (N Chomsky, 1957). However, the neuroscientific field has shown how the infant brain develops alongside language learning in the first few years of life challenging the notion that the language system is ready for language at birth. Instead, the infant brain changes considerably during the first few years of life as language abilities improve (Friederici, 2017b). Infants and toddlers at a very early age are exposed to speech sounds and different sequences of sensory events as they learn language. They also advance their acquisition of language and the learning of language patterns by hearing their parents and other individuals talking. For instance, infants can extract information from sequences of speech sounds by using the statistical relationships between speech sounds or made-up nonsense words (Saffran, Aslin, & Newport, 1996). Infant studies show that the developing human brain's sensitivity to specific sound transitions is related to the ability to detect sound temporal patterns in time (Jusczyk & Aslin, 1995).

Regarding multisensory integration, it was found that infants could understand the meaning of different words in combination with visual objects, being able to associate sounds with visual objects (Bergelson & Swingley, 2012; Schacter, Gilbert, Wegner, & Hood, 2011). These findings indicate that humans at very early age can comprehend the meaning of speech sounds, integrate audio-visual events and interpret relationships between elements in sequences, some of which have been directly associated with their language learning capabilities. Unlike humans, monkeys do not have human language, but as noted, are nonetheless able to demonstrate referential vocalisation, such as producing predator alarm calls (R. M. Seyfarth et al., 1980), and they are capable of learning to notice the temporal regularities of sounds in a sequence like humans (Wilson et al., 2015). However, monkeys in the wild may not be often exposed to human language-like environments, and thus they may not have the chance to experience complex sounds in similar ways as humans in their natural environment. That said, listening to complex sounds does not necessarily lead to comprehension of those sounds, as is the case for listening to a foreign language that humans are unfamiliar with. Therefore, it is worth noting, that comparisons to humans are often not direct for these many reasons, including monkeys are trained for many months to conduct tasks that are easier to implement in humans often within an hour simply by instructing the human participants using language (Poletiek, Monaghan, van de Velde, & Bocanegra, 2021). Therefore, it is important to keep track of the divergence in how the species are studied, such as the amount of time the animals require for progressing through a task. Scientists are mindful that comparisons with humans will not always be direct even if the task conducted in both species is identical.

1.2 Animal Models for Comparative Cognition and Studying Language Precursors via Touch Screens

In this section, I overview the prior neurobiological literature on language precursors in NHPs.

Animal models are critical for a better understanding of neurobiological function and treatment for disorders of brain, particularly the use of NHPs as they share a recent common ancestor with humans. Old World Monkeys separated from humans about 25 million years ago, and New World Monkeys about 50 million years ago, meaning that Old World Monkeys are evolutionarily more closely related to humans.

Nonhuman primates play vital roles with regard to cognition as well as translational research from animal models to humans (Friedman et al., 2017). Macaque monkeys have been used widely in the laboratory as an important neurobiological animal model due to their anatomical and physiological cortical similarities to humans (Phillips et al., 2014), making them an important neurobiological model system for humans (Mason et al., 2019). The issue, however, as this thesis aims to address is how realistic of a model macaque monkeys are regarding key properties of language.

Brain regions such as prefrontal cortex (PFC) in macaques are functionally similar to humans with regard to decision-making and attention (Haile, Bohon, Romero, & Conway, 2019). Importantly, the amygdala and some subdivisions of PFC make critical contributions to reward-associated tasks and emotional learning in both macaque and human brains (Bermudez & Schultz, 2010; Gottfried, O'Doherty, & Dolan, 2002; Kaskan et al., 2016; Li, Schiller, Schoenbaum, Phelps, & Daw, 2011; O'Doherty, Rolls, Francis, Bowtell, & McGlone, 2001; Paton, Belova, Morrison, & Salzman, 2006). This provides a basis that macaque monkeys would be an important animal model for neuroscience studies, yet a similar argument could be made for other species more distantly related to humans (e.g., New World Monkeys). I therefore as part of the planning for this thesis work, needed to choose which species to study given that Newcastle University has both macaque and marmoset monkeys that could have been tested. My decision was based on the following reasoning.

The basic combination of different calls with specific meanings was found in Old World Monkeys by combining two distinguished vocalisations into a short sequence. The monkeys were able to generate a predator alarm call combination that led to a different behaviour than the usual response to the alarm calls individually (Arnold & Zuberbühler, 2006). Much less is known about the extent of New World Monkeys' combinatorial calling capabilities. It is also known that macaque monkeys share similarities with humans in their auditory and visual pathways (Balezeau et al., 2020; DeValois & Jacobs, 1971; Stebbins, 1971). Again, much less is known in these regards for New World Monkeys (Wilson et al., 2013).

Apes are proficient in symbolic learning and new research has shown that they use combinations of vocalisations to expand their communication capabilities (Berthet, Surbeck, & Townsend, 2025; Lameira, Caneco, Kershenbaum, Santamaría-Bonfil, & Call, 2025). And apes, as a specially protected species, can be studied behaviourally but not neurobiologically. Although Newcastle University had marmosets and macaques as the leading animal model species throughout this PhD thesis work, I decided to focus on macaques as a species more closely evolutionarily related to humans for the reasons noted above. Based on the observations obtained here with macaques, future work could be conducted with touch screens using similar tasks in marmosets or apes.

Studies have demonstrated that humans and NHPs share similarities in terms of neural mechanisms supporting cognition more generally, including decision-making and information-integration for complex cognition (Bondar, Vasileva, Tereshchenko, Buynevich, & Latanov, 2019). Moreover, studies such as sequence learning involved in language processing indicated that NHPs can perform sequence learning tasks with similar patterns as humans (Conway & Christiansen, 2001; DeLillo, 1996; Terrace & McGonigle, 1994), and the general neural mechanism in sequence prediction with relation to language might be evolutionarily conserved in NHPs (Kikuchi, Sedley, Griffiths, & Petkov, 2018). Additionally, the neural mechanisms for recognition of faces and voices as well as the integration of them in NHPs are shared with humans (P Belin & Zatorre, 2003; Freiwald, Tsao, & Livingstone, 2009; Ghazanfar & Eliades, 2014; Kanwisher, McDermott, & Chun, 2002; Petkov et al., 2008). Another aspect of human language is the capability of vocal production learning, whereby infants begin to babble and then speak the language of their parents and community. Studies have shown that nonhuman primates are capable of learning vocalisations and producing natural vocalisations in different environmental contexts (Hage, 2018; Lattenkamp & Vernes, 2018; R. Seyfarth & Cheney, 2018), which might suggest that monkeys can learn the meaning of auditory stimuli if they are given access to learn specific meanings when conducting the behavioural tasks.

In order to have a better understanding of the mechanisms of perception and cognition, of relevance for modelling human neurobiology using approaches not feasible for study in humans, monkeys are important to be used as an animal model. However, the uncertainty on how good of a model the nonhuman primate brain could be for humans requires further understanding the extent of their cognitive abilities and the links to language in humans. One area in primate cognition that is not explored is whether NHPs can combine information from one sensory modality (auditory) to identify objects in another (visual), as well as their still undefined limits of learning abilities. To better understand primate cognition, behavioural testing can be conducted to understand the complexity of primate cognition. In addition, these approaches allow close comparisons between NHPs and humans. More specifically, visual stimuli can become associated in long-term memory when they are frequently seen together, and auditory stimuli can become associated when they are frequently heard together. In other words, nonhuman primates might be capable of learning the associations between auditory and visual stimuli with greater structural complexity in a sequence (Rey, Perruchet, & Fagot, 2012).

Touchscreen training can be applied to evaluate animal cognitive functions particularly on complex tasks, opening new doors in understanding the extent of their abilities and which aspects of human cognition they can model. There has been a widely used system developed by the University of Cambridge for home-cage testing through a touchscreen named Cambridge Neuropsychological Test Automated Battery (CANTAB). The advantage of this system is that it saves the technicians as well as the researchers a lot of time to train the animals as it can be attached to the home-cage with no need for people to be around. Apart from that, it can be applied not only to macaques, but also to marmosets (Robbins et al., 1994). The CANTAB system plays an important role with regards to the assessment of the visual and spatial learning, as well as the memory (Robbins et al., 1994).

Other developments in this aspect include a wireless touchscreen system for training NHPs that was established by the University College London and named Mymou (Butler & Kennerley, 2019). The advantages of this system are that some problems during the touchscreen task can be overcome such as separation of animals including restraints and limited collection of data. Furthermore, there have also been some other developments such as the apparatus with a touch-sensitive screen for evaluating cognitive functions in marmosets, utilising a laptop PC combined with a touch-sensitive screen as the major components of the system revealed the efficacy by testing visual discrimination and reversal tasks to promise researchers an easy way proceeding daily experiments (Takemoto, Izumi, Miwa, & Nakamura, 2011). Therefore, touch screen training can be applied to animal learning particularly on complex tasks, opening new doors in understanding the extent of their abilities and which aspects of human cognition they can model (Mitchell, Reynolds, & Miller, 2014; Roberts et al., 1990; Yamazaki et al., 2011). As another example, the study by Mitchell (2014) suggests that marmosets are a good animal model for studying neural mechanisms, and the application of various technologies in behavioural conditions, showing a reliable performance for easy visual discriminations for cognitive functions.

Nevertheless, macaques might be the best model due to their larger brain size and shared similarities of the pathways as well as regions of the brain with humans, by which make them a well-established animal model for comparative studies (Baars & Gage, 2010; Bitar, Mansour, & Chehab, 2016; Perrodin, Kayser, Abel, Logothetis, & Petkov, 2015).

1.3 Combinatorial Learning Using Touchscreen Tasks and the Need for Data on Primate Models

Given that some of the basic semantic and syntactic functions are evolutionarily conserved, an important question is whether NHPs can integrate disparate sensory information into what may be a basic combinatorial form. For example, Noam Chomsky has argued that a very basic Merge operation is foundational to language (Noam Chomsky, 2020). This Merge operation could at this basic level combine two pieces of information (A and B) into a phrase that carries joint meaning that can be used to identify an object by multiple features ('big dog' etc.). Chomsky argues that this basic Merge operation is used repeatedly to create sentences with more complex communicative content. Although Chomsky considers this Merge operation to be human unique, without direct evidence this basic Merge operation however might be evolutionarily conserved.

There was a human study conducted by Poletiek et al, who used both audio and visual stimuli to explore language learning abilities in humans (Poletiek et al., 2021). The paradigm is the closest correspondence of the experiments I tested in animals, and I collaborated with Prof. Poletiek on this work. The paradigm was designed into 4 experiments, and the participants were tested whether they were able to learn sequences of combinations of different colours and shapes followed by the structure, and the learning of combinatorial dependencies was required. For instance, a series of A(s) being placed to the left of a B following the structure of A^nB^n, and were presented in the way of A(1) A(2) A(3) B(3) B(2) B(1). These A and B letters here represent different meanings such as the shape square or the colour yellow, respectively. Also, these AB pairs could be formed into sentences with different levels of embeddings, started with 0 level of embedding, which was simple AB pairs, followed by the first level of embedding in the structure of AABB sequences. Finally, the AAABBB structure of the 2 level of embedding, which represented 3 objects. By presenting these coloured shapes with associated sounds, different sequences could be generated hierarchically. In the experiments, human participants took a short time to learn the associations between the auditory and visual stimuli, and they were able to learn the artificial grammar language across modalities and to the level of more complex, more language-like hierarchical dependencies. Moreover, they were able to generalise to novel sentences not previously learned by using the basic learned combinations of semantic elements and their grammatical structure (syntax).

A recent study by Medam and Fagot was a first attempt to identify if monkeys could integrate visual objects in a basic combinatorial way. The authors evaluated whether baboons could identify a complex visual symbol that referred to an object by both its colour and shape features (Medam & Fagot, 2016). They conducted the study asking whether baboons working for reward on touch screens can understand an abstract object (the letters 'ND') on the screen and use them to identify an object that has the combination of the shape and colour features indicated by the labels (for instance, N equals red, and D equals cross). More specifically, the baboons were presented with English letters (e.g., 'N', 'D', etc.) which were associated with colours and shapes (e.g., 'N' equals red, 'D' equals cross). Additionally, the baboons were presented with a composite stimulus (e.g., 'ND'), alongside the coloured shape that these stimuli represented. They were also presented with several foils (different coloured shapes), and they received a reward only when they correctly selected the shape that matched the letters (in this case, the red cross) (Medam & Fagot, 2016). The baboon would need to reject several foils that have objects that might share one of the features or none.

In two experiments, the baboons were trained with either the paired symbols or individual symbols at a time, in each case needing to select the correct shape and colour objects. In the first experiment, the baboons were trained directly with the combined labelled symbol (e.g., ND), and in the second experiment, the animals were trained with individual colour (e.g., N) or shape (e.g., D) labels. In both experiments, the baboons needed to select an object that had the combined or one of the features. Notably, all the baboons (all female) had 24/7 access to the touchscreen, allowing them to perform the task at any time. In addition, a considerable number of training trials were needed for the baboons to meet the set criterion for learning the basic object features, suggesting that the task was more difficult than expected and required significant effort for any of the baboons to meet the criterion.

Only one of the baboons in the first experiment was able to successfully identify the joint object features, suggesting that the combined features are difficult to immediately extract by most baboons. Therefore, in the second experiment, the authors attempted to train the baboons on the individual shape and colour features before these were combined. However, even in the second experiment, none of the 11 baboons involved in this experiment were capable of showing significant learning. The authors considered several aspects of the experimental design that may have led to the baboons not being able to demonstrate strong combinatorial semantics learning. There were also different sample sizes in the two experiments, for reasons that are unclear. For instance, there were 7 baboons in experiment 1, and another 4 in experiment 2 including the subset from experiment 1. My hypothesis on the experiment is that even the second experiment was very difficult for the monkeys, because even the single symbol on the screen (e.g., N or D) required selecting the features from objects that also had other features (a correct answer could be a red cross, a red circle or red anything). In other words, the letter 'N' means red, and the stimuli on the screen are red cross, blue diamond and yellow cross, and the monkey needs to figure out not only which stimulus is correct, but which feature of the stimulus they should be paying attention to. Therefore, even in the 'simpler' experiment 2, the monkeys may have gotten very confused on their task.

Therefore, I think that a key issue is that the baboons involved in both of their experiments were not trained as simply as could have been done, step by step. In other words, the baboons were not taught or trained the associations between the labels and the object features in a way that could have been simpler. The symbols were presented on the screen either in an already combined two-symbol fashion or even when they were presented individually, the baboons needed to figure out the colour or object amongst many objects, which likely caused substantial confusion even for the one baboon that appeared to figure out the combinatorial properties of the task in the first experiment. Therefore, these experimental decisions might have increased the difficulty of the task itself for the baboons. Although the baboons had unlimited access to the touchscreen as mentioned previously, this might not have been helpful to them in terms of the clarity of the task that they were asked to do. Moreover, the experiments were conducted in the visual modality only, providing few insights into whether the monkeys could use one modality to identify objects in another, as an evolutionary precursor process for language would need to be to flexibly apply a cognitive process across sensory modalities. Moreover, the auditory modality was likely the precursor system of choice for speech processing (Balezeau et al., 2020; Rocchi et al., 2021).

The auditory modality is important to study because nonhuman animals can produce vocalisations (Levelt, 1999; Sapir, 1921). The production of speech is closely linked to the organs involved such as the tongue and lips which are more related to physiology and functional areas in the brain (Sapir, 1921), and the acquisition of speech is the process of integration and understanding sounds produced by other humans by which human listeners are able to comprehend via language.

Therefore, the pioneering work by Medam and Fagot, and the need to study the auditory modality sets the stage for my proposed work here. For me, it was important to understand the limitations of the study in engaging the baboons to perform better with the combinatorial aspects of their task. Moreover, I wanted to apply the approach to the auditory modality. For example, the research conducted by Seyfarth et al. (R. M. Seyfarth et al., 1980), studying monkeys' responses to different alarm calls gives initial confidence that if the monkeys can classify several predators, then auditory semantic capabilities among the monkeys may be possible whereby the laboratory macaques may be able to associate different sounds to different visual object features.

Specific Hypotheses:

- I hypothesise that macaque monkeys are able to use a short sequence of sounds to identify visual objects by colour or shape.
- Furthermore, I hypothesise that although this type of auditory to visual associative learning will be difficult for the monkeys, that they may be able to learn through trial and error (with reward for correct trials) that a two-sound sequence of sounds refers to an object by both its colour and shape features. To do this they must demonstrate that they can reject incorrect foil stimuli.
- I also predict that the monkeys may be able to generalise their training to combinations of sounds that refer to objects for which they are not trained on, which I will assess by using probe trials that are not rewarded, presenting the combinations that the monkeys are not trained on.

1.4 fMRI Study Overview of Sequence Learning in Nonhuman Primates

Although I was able to complete the data collection for the audio-visual functional magnetic resonance imaging (fMRI) study in macaques and humans using other types of Artificial Grammar Learning sequences, the demands of the combinatorial learning task with the macaques prevented me from completing the fMRI data analysis and study reporting. Therefore, I here provide a brief overview of some of the neuroscientific information available on how the primate brain processes the stimuli that I will be presenting to the monkeys, with some consideration of the brain areas that may be involved in this form of complex cognition.

As aforementioned, semantics and syntax play vital roles in language, and NHPs share remarkable similarities with humans concerning the basic aspects of language functions (e.g., semantics and simple syntax-like constructions). Therefore, the neurobiological study on NHPs may provide insights into language processing such as semantics, syntax and their interplay. Studies suggest multisensory communications among nonhuman animals such as NHPs is a common feature of animal cognition (Ghazanfar & Schroeder, 2006; Ghazanfar & Takahashi, 2014; Jordan, Brannon, Logothetis, & Ghazanfar, 2005; Leavens, Russell, & Hopkins, 2010).

Sequence learning abilities are viewed as syntactic abilities of a potential precursor to human language (Friederici, 2018; Gervain, Berent, & Werker, 2012; Gómez & Maye, 2005; Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Milne, Petkov, & Wilson, 2018; Mueller, Friederici, & Männel, 2012; Petkov & Wilson, 2012). Studies indicate that sequence learning involved in language processing indicated that NHPs can perform sequence learning tasks with similar sensitivity to humans at least for adjacent sequencing dependencies (Conway & Christiansen, 2001; DeLillo, 1996; Terrace & McGonigle, 1994). Therefore, the general neural mechanisms in sequence prediction may have been an evolutionary precursor substrate present in nonhuman primates, upon which human language evolved (Kikuchi, Sedley, Griffiths, & Petkov, 2018). However, not much is known about how the brain processes the sequences that have the same structure across modalities in comparison across different species.

Perisylvian language network supports human language, and they consist of a range of regions and subregions around the sylvian fissure, which separates the temporal lobe and dorsal frontal as well as frontal parietal cortices (Catani, Jones, & Ffytche, 2005). Broca's area and Wernicke's area form the foundation of the language network: Broca's areas (Brodmann's area 44 and 45) are found in the left inferior frontal gyrus, and connected through the arcuate fasciculus to Wernicke's area, which is found in the posterior temporal areas (superior temporal gyrus) and correspond to Brodmann's area 22 and 42 (Aboitiz & Garcıa, 1997; Catani et al., 2005; Dejerine & Dejerine, 1895; Penfield & Roberts, 2014; Wise et al., 2001). This network has now been extended from the originally defined language regions, and several pathways are demonstrated to support language processing. Two dorsal pathways start from posterior superior temporal gyrus (STG), and one of them interconnects with the premotor cortex, which is implicated in integrating auditory information relevant for vocal production. Another dorsal pathway travels through the arcuate fasciculus and terminates at Brodmann's area 44, which in humans is functionally involved in complex syntactic processing (Friederici, 2004, 2011, 2012, 2017a; Friederici, Mueller, & Oberecker, 2011; Liu et al., 2023). Ventral pathways are connected with more anterior areas and middle superior temporal gyrus to the more ventral areas of inferior frontal gyrus, which are thought to support semantic and syntactic processing, and one of the ventral pathways connects superior temporal gyrus to the frontal operculum (FOP) via the uncinate fasciculus (Friederici, 2004, 2011; Friederici et al., 2011; Perani et al., 2011; Rilling et al., 2008). These suggest that simple syntactic processing is supported by a ventral pathway from STG to FOP, and the complex semantic and syntactic processing is supported by a dorsal pathway from STG to BA44/45. A recent structural connectivity study in humans, apes and macaques in the laboratory has shown that macaques possess a dorsal pathway interconnecting inferior frontal with posterior temporal lobe auditory areas (Balezeau et al., 2020).

A number of brain regions in NHPs share similarities structurally and functionally with human brain homology in relation to the perisylvian language network such as the area 44, 45 in prefrontal cortex (Petrides & Pandya, 1999, 2002). Connectivity between 44d or 44v (subregions of BA44, which are antero-dorsal and ventro-posterior parts) and areas of 45A and 47/12 is also found to be similar between monkeys and humans (Neubert, Mars, Thomas, Sallet, & Rushworth, 2014; Zilles & Amunts, 2018). There are differences, however, the interconnections and density of the dorsal pathway in macaques are found to be sparser than in humans, while the ventral pathway is larger in monkeys and apes relative to adult humans (Friederici, 2017a; Rilling et al., 2008).

Cognitive neuroscience studies the biological processes and aspects underlying cognitive functions. Since monkeys do not have human language, one might look for precursors to the language system within the cognitive system of NHPs (Gazzaniga, 2009). Complex cognition refers to cognitive functions that require integrating information across multiple levels. For example conditions with regards to the mental functions that require acquiring information in one modality and applying it to guide behaviour in another, a process that requires perception, learning and memory (Prinz, Aschersleben, & Koch, 2009). Although monkeys do not have language, there is evidence from other domains of cognition whereby macaques were able to integrate different facets of information, suggesting that their cognitive system is even more sophisticated than previously thought (Ferrigno, Cheyette, Piantadosi, & Cantlon, 2020).

Regarding the auditory and visual processing in the brain, the temporal lobe plays an important role in both auditory and visual processing (Perrodin et al., 2015). Additionally, the superior temporal sulcus (STS) is thought to contain conceptual representations of semantic knowledge at least in humans and is involved in sensory convergence in both humans and NHPs. It is involved in auditory input processing, reliant on the auditory cortex located on the upper banks of the temporal lobe as well as the cortex below and within the Sylvian fissure (Baars & Gage, 2010). The temporal lobe plays an important role in processing sound, specifically, the primary auditory cortex and the superior temporal gyrus are involved in the initial stages of sound processing in cortex (Patel, Biso, & Fowler, 2022). Regions in the temporal lobe that are more inferior and posterior are involved in recognising visual objects, such as the human fusiform gyrus or the inferior temporal (IT) cortex in macaques (Baars & Gage, 2010). As in macaques and humans, it is now known that two streams consist of the primary visual pathways which are named the ventral stream and dorsal stream (Bitar et al., 2016). The ventral stream originates in primary visual cortex (V1), going through the visual area 2 (V2) and the extrastriate area V4, and to the inferior temporal area (IT), while the dorsal stream starts from V1 as well, then going through the dorsalmedial area (V6) and medial temporal area (MT), and to the posterior parietal cortex (Ungerleider, 1982). The ventral stream of visual processing is involved in object perception and recognition, and the temporal lobe has a close connectivity with the hippocampus and amygdala for long-term memory storage and concept formation (Schacter et al., 2011). The superior temporal sulcus (STS) extends along the lateral side of the temporal lobe (Lanfermann, Raab, Kretschmann, & Weinrich, 2019), separating the superior temporal gyrus from the middle temporal gyrus in the temporal lobe or the IT cortex in macaques. Studies revealed that the STS is a region of multisensory convergence with audio/visual stimuli, containing a significant interconnectivity between visual, auditory and somatosensory areas (Kropotov, 2010). A human fMRI study indicated that neural representations of audio-visual integration, vocal sensitivity, and facial sensitivity were all located in different areas of the superior temporal sulcus (Kreifelts, Ethofer, Shiozawa, Grodd, & Wildgruber, 2009). This means that the STS is a brain region for sensory convergence, receiving multisensory input and playing a crucial role in perception and cognition based on information from all sensory inputs (Specht & Wigglesworth, 2018). The superior temporal gyrus (STG), just above it, can also be influenced by multisensory information presumably because of interactions with the STS (Binder, Hirokawa, & Windhorst, 2009) although this needs to be better understood at the neuronal circuit level. Visual-sensitive neurons were identified in the area of inferior temporal (IT) in macaques, and these neurons have a stronger response to visual objects (Bruce, Desimone, & Gross, 1981; Perrett, Rolls, & Caan, 1982). Additionally, studies by neuroimaging indicated that neurons in the inferior bank of the superior temporal sulcus (STS) respond more strongly to visual stimulus in monkeys (Perrodin et al., 2015). While in humans, regions that are more sensitive to visual input are in the fusiform gyrus (FG) and occipital areas (Perrodin et al., 2015). As for sensory convergence, the upper bank of the superior temporal sulcus (STS) is considered to be a key site for multisensory convergence, specifically, the fundus and the lower bank of the STS potentially contain visual-sensitive clusters (Perrodin et al., 2015).

Moreover, with regards to colour-sensitive regions in NHPs, studies on macaques indicate that V1 and V2 of visual cortex have the exhibition of functional domains for processing colour vision (Salzmann, Bartels, Logothetis, & Schüz, 2012). Additionally, V4 as one of the extrastriate visual areas located anterior to V2 and posterior to the posterior inferotemporal area (PIT) has strong connections with V1 and V5, and is closely involved in colour and shape perception (Tanigawa, Lu, & Roe, 2010) in macaques.

It is also not clear which part of the auditory cortex would be most likely to influence sensory convergence in the STS. For instance, the auditory cortex can be subdivided into the core, belt, and parabelt which are areas including primary auditory cortex (A1), the secondary auditory cortex (A2), and the tertiary auditory cortex (A3) respectively, and the belt surrounds the core immediately (Pickles, 2013). The anterior lateral area (AL) is located in the portion of the auditory lateral belt that neurons are more sensitive to broadband noise and complex sounds than that neurons in the core (Brosch, Schulz, & Scheich, 1999; Hackett, Preuss, & Kaas, 2001; Heil, Scheich, Budinger, & Konig, 2005; Kadia & Wang, 2003; Kikuchi, Horwitz, Mishkin, & Rauschecker, 2014). In parallel, the middle lateral area (ML) is located in the portion of the belt that lateral to the primary auditory cortex, and neurons in ML as well as adjacent areas are more responsive to band-pass noise bursts with defined centre frequencies than tones (Rauschecker, Tian, & Hauser, 1995). Apart from that, the rostro-temporal area is located in portion of the core and rostral to the parabelt, additionally, the medial rostro-temporal has a weak connection with part of the superior temporal gyrus (STG) (Heil et al., 2005). Auditory-sensitive areas in monkeys were identified in the anterior superior temporal gyrus (aSTG), temporal polar cortex, anterior insula, and ventrolateral prefrontal cortex (vlPFC) (Perrodin et al., 2015). In correspondence, auditory-sensitive areas in humans such as STS and STG (Pascal Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) have been studied using functional magnetic resonance imaging (fMRI).

The primate brain is also known to be highly multisensory. For example, there are auditory inputs into visual cortex and vice versa, and some have argued that there is no 'unisensory' cortex anywhere in the brain (Ghazanfar & Schroeder, 2006). However, this might apply for basic auditory/visual properties (like when a sound/object needs to be identified very quickly). Frost et al., however, argue that when it comes to integrating information over time (sequential information) the sensory systems are surprisingly separate (Frost, Armstrong, Siegelman, & Christiansen, 2015). This makes a very interesting prediction about how the primate brain processes auditory and visual sequences that have the same structure over time, as might be the case for the behavioural experiments that form this thesis research.

This background of work forms the specific hypothesis of the neurobiology study as flows:

1) Some parts of the superior temporal sulcus (STS) and the prefrontal cortex will process the structure of the sequence, regardless of whether they are auditory or visual, showing very similar fMRI patterns to the auditory and visual sequences obtained in separate experiments and testing sessions.

2) Alternatively, different patterns of activity may occur for auditory and visual sequences, meaning that these areas in the brain are modality specific. This would be consistent with the model Frost and colleagues (2015), who proposed that the sensory systems remain separate when it comes to integrating the structure of auditory and visual sequential information over time.

To test how the primate brain processes auditory and visual sequences that have the same structure over time, we have designed an experiment to conduct with human and macaque fMRI. I was able to help complete the data collection and show some initial results in Chapter 6 of this thesis.

Chapter 2: Developing Effective Touchscreen Training Methods: Task Design and Training Strategy for Rhesus Macaques

Parts of this chapter have been presented as an oral presentation at the following scientific conference

Zhang. Z., Wilson. B., Muers. R., Milne. A., Kikuchi. Y & Petkov. C. (2022) Complexity of Primate Cognition and Neural Bases of Sensory Convergence. North East Postgraduate Conference, Newcastle, 2022.

Abstract

How human speech and language evolved from an auditory system shared with ancestors to living nonhuman animals remains an important open question, one with implications for the extent to which aspects of the human language system can be modelled in nonhuman animals. A key property of human language is combinatorial semantics, where information from a sequence of words is integrated to identify meaningful content. We designed a novel behavioural touchscreen task implemented with two rhesus monkeys (*Macaca mulatta*) in their home units. The task allowed us to study whether the monkeys could associate nonsense speech sounds with visual colours or shapes, prior to integrating the information contained in a sequence of two sounds identifying a specific object by its joint colour and shape properties. In this Chapter, the touchscreen training methods and the task will be described.

2.1 Introduction

The world is multisensory, yet how the brain integrates sensory information is not well understood. Identification or recognition can be achieved across modalities, for instance, objects often have visual/auditory or other sensory components. It is known that for basic sensory features, the brain is highly multisensory (pathways between auditory and visual cortex, multisensory influences throughout)(Ghazanfar & Schroeder, 2006; Perrodin et al., 2015). Studies on humans in relation to the perception of face-identity and voice-identity suggested that the convergence of auditory and visual sensory in anterior temporal lobe (ATL) (Blank, Wieland, & von Kriegstein, 2014; Campanella & Belin, 2007; Schweinberger & Burton, 2003). Yet when sounds and pictures have structured information over time, some have argued the sensory systems stay segregated (i.e., there is not a common temporal structure applied to both sounds/pictures) (Frost et al., 2015). Thus, it is important to understand the complexity of primate multisensory integration abilities.

For instance, work in other domains of neuroscience suggests that primates in particular could integrate multiple forms of information. Recent work has shown that primates (like macaques) can integrate different types of information: reward magnitude and probability (Bongioanni et al., 2021), integrate the length of a sequence of sounds with what type of sequence it is (Wang et al., 2019), integrate information about multiple visual object dimensions (Medam & Fagot, 2016).

Understanding the extent of the combinatorial learning capabilities of NHPs is an outstanding scientific question. Using a novel combinatorial learning task developed in humans (Poletiek et al., 2021), I aim to determine whether monkeys trained to associate speech sounds to visual object colour or shape can combine these to hear two sounds that determine an object by both its shape and colour. Determining whether the monkeys can integrate auditory information to identify both object colour and shape would be a substantial discovery. It is also possible that they might only be able to choose colour or shape but not both, if they are not able to integrate both features but are able to rely on one or the other. A prior study (Medam & Fagot, 2016) showed that at least one monkey can learn a visual combinatorial learning task similar to the paradigm that I am using. However, that study did not step the monkeys through the different stages of learning very well, so it is possible that many more monkeys are capable of combinatorial learning. As mentioned, their baboons had free access to the touchscreen 24 hours/7 days which might help in increasing the number of trials that the baboons can perform as the testing time was not limited. From our side, the monkeys are tested in the home-cage between 2-3 hours each working days.

This basis forms the hypotheses that will be tested: 1) that macaque monkeys can learn to associate a sound with either colour or shape; 2) that macaque monkeys can learn to identify an object by its shape and colour solely by hearing a short sequence of two sounds that identify those features; 3) that macaque monkeys can do both shape and colour associative learning (mapping a sound to each object feature) but struggle to integrate them, always preferring to identify either colour or shape, or to shift between identifying one or the other but not fully integrating them. A strong test of this hypothesis is testing the animals with combinations that they would not have experienced during training, as I describe further below to see if they generalise the learning to unexperienced combinations.

2.2 Touchscreen Training Methods

2.2.1 Subjects and Housing

We tested two rhesus macaques (*Macaca mulatta*) on the touchscreen task, one male and one female, and they were housed in the Nonhuman Primate (NHP) area of the Comparative Biology Centre (CBC), Medical School, Newcastle University, the UK under Home Office licensing and regulation (Animals (Scientific Procedures) Act 1986), including project license and personal licenses for all the personnel working with the primates. The male (age: 13, weight: 14 kg, in-pair with another female formerly and parried with the female involved in the touchscreen task later) and the female (age: 7, weight: 8.5 kg, housed individually before, and in-pair with the male in the touchscreen task later) were tested in their home units between 2 and 3 hours five days a week in the mornings or afternoons. These two macaques were housed in-pair in the same home-cage later, and they were separated in different cages while being tested.

2.2.2 Touchscreen Testing Apparatus

Figure 2.1. illustrates the apparatus of the touchscreen device. The touchscreen (15-inch ElO LCD touchscreen, Model: ET1590L) is attached to the home cage during the testing with the frame stopping the monkeys from getting access to the other parts of the device excluding the touchscreen itself. And the laptop (Dell Latitude 7490) is installed with Matlab (Matlab R2017a)

and Psychtoolbox 3 (Psychophysics Toolbox Version 3) to run the experiment. All components are connected through cables which kept out of reach of the macaque’s grasp. The laptop and system were battery (SUNGZU, Model: SKD200-1) powered for safety reasons to avoid running a power cable from a power socket to the setup. The audio-speakers (Manhattan 2600 Series Speaker System) are installed inside the lower box of the touchscreen device and connected to the laptop to allow the sound being clearly played and heard by the monkeys. The speed controller (Audon Electronics) can be adjusted to control the speed of juice reward when being given to the monkeys and it is set before the testing started. Additionally, the juice comes through the juice tube if the monkey completes a correct trial. After the testing session, the juice tube and the touchscreen device are cleaned and put back to the storage room for charging.

Touchscreen Apparatus

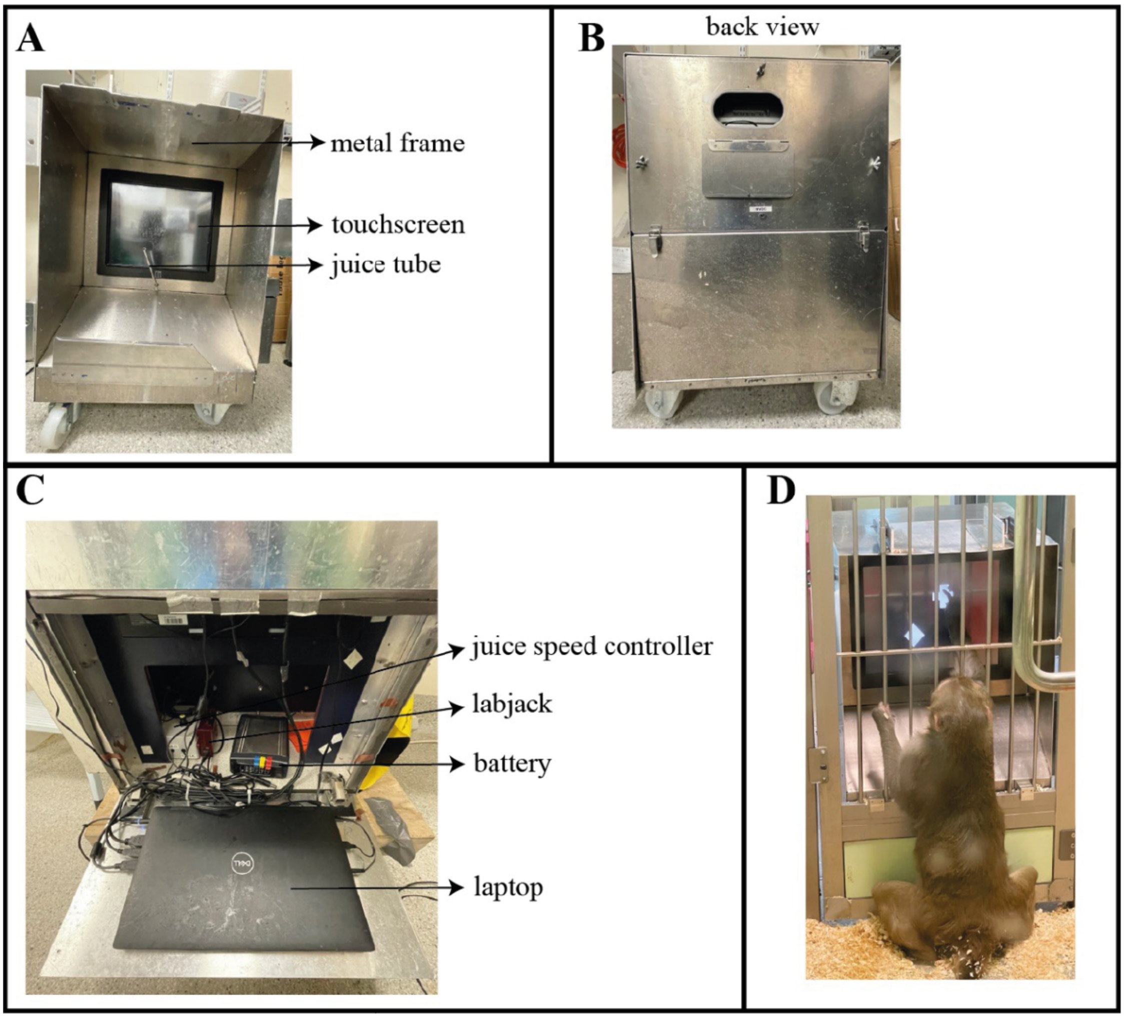


Figure 2.1. Touchscreen apparatus. Panel A represents the front view of the touchscreen apparatus which includes a metal frame, the touchscreen and the juice tube. Panel B is the back

view of the touchscreen and the cables connected to the laptop as well as other devices are locked in the box at the lower bottom. Panel C indicates some parts of the important devices such as the juice speed controller, the Labjack (Audon Electronics, Model: U3-LV), the external battery, and the laptop. Panel D shows an example session that the monkey performs the touchscreen task.

2.2.3 Stimuli

The auditory stimuli were three computer-generated acoustical sounds using Audacity modified at 48000 Hertz, and the sound duration of each auditory stimulus was 0.6 seconds. In parallel, the visual stimuli were composed of three different shapes and three different colours with apparent discriminability (200 * 200 pixels each), and the combined shapes and colours were the integrations from the colours and shapes. These visual stimuli were drawn from Inkscape and presented on the black background, and the colours appeared to be blurred to avoid any possibility of being represented as a ‘shape’.

2.3 Touchscreen Task and Procedures

Pilot data was collected in M1 before the start of the touch screen training, and M2 was experienced with the lab training before the touch screen training as well. Specifically, the pilot data was collected by one of the former colleagues in the lab, and the task required the monkey to make a choice between different objects which were the pictures of an orange cat, a brown dog and a black-and-white rabbit that had the same pixels. Additionally, sounds associated with the visual objects were played before the visual stimuli were presented on the screen. The monkey needed to make a choice between the objects, and the reward was given if the monkey made a correct choice. We were confident that the monkeys could recognise these visual objects because macaque and human visual perception is highly similar. For instance, visual colour can be used to guide response inhibition in macaque monkeys, suggesting that colours could be used as cues in cognitive studies in both species (Blizzard, Fierro-Rojas, & Fallah, 2017; Ghasemian, Vardanjani, Sheibani, & Mansouri, 2021a).

In our case, we designed a novel behavioural touchscreen task implemented with two rhesus monkeys (*Macaca mulatta*) in their home units. The task allowed us to study whether the monkeys could associate the nonsense speech sounds with the colours or shapes, prior to integrating information from a sequence of two sounds identifying a specific object by its joint colour and shape properties. The paradigm was implemented in three key phases as can be seen in Figure 2.2. B. In the first phase, the animals started by learning to associate acoustically presented nonsense words with specific colours or shapes. In the second phase, the shape and colour versions of the touchscreen learning were alternated week by week, day by day or within the same day. This means that the monkeys were tested with the shape or colour only tasks during the alternation, with alternation being slower or quicker with different rates depending on their performance. They were tested with either more colour or shape versions depending on whether they showed any bias towards the shape or colour property. Two sessions of the behavioural data were collected when the monkeys were tested with the alternation of colour and shape versions within the same day. When the monkeys reached out the threshold which was defined as a majority 6 out of 10 'good' sessions within two weeks of testing (5 days a week), it was a good indication of moving the monkeys to the next testing phase. Notably, the sessions here referred to the days as two runs were tested within the same day on colour and shape versions. Good performance is scientifically defined as can be found, e.g., Figure. 3.11, where the monkey reached the significance threshold on a given session.

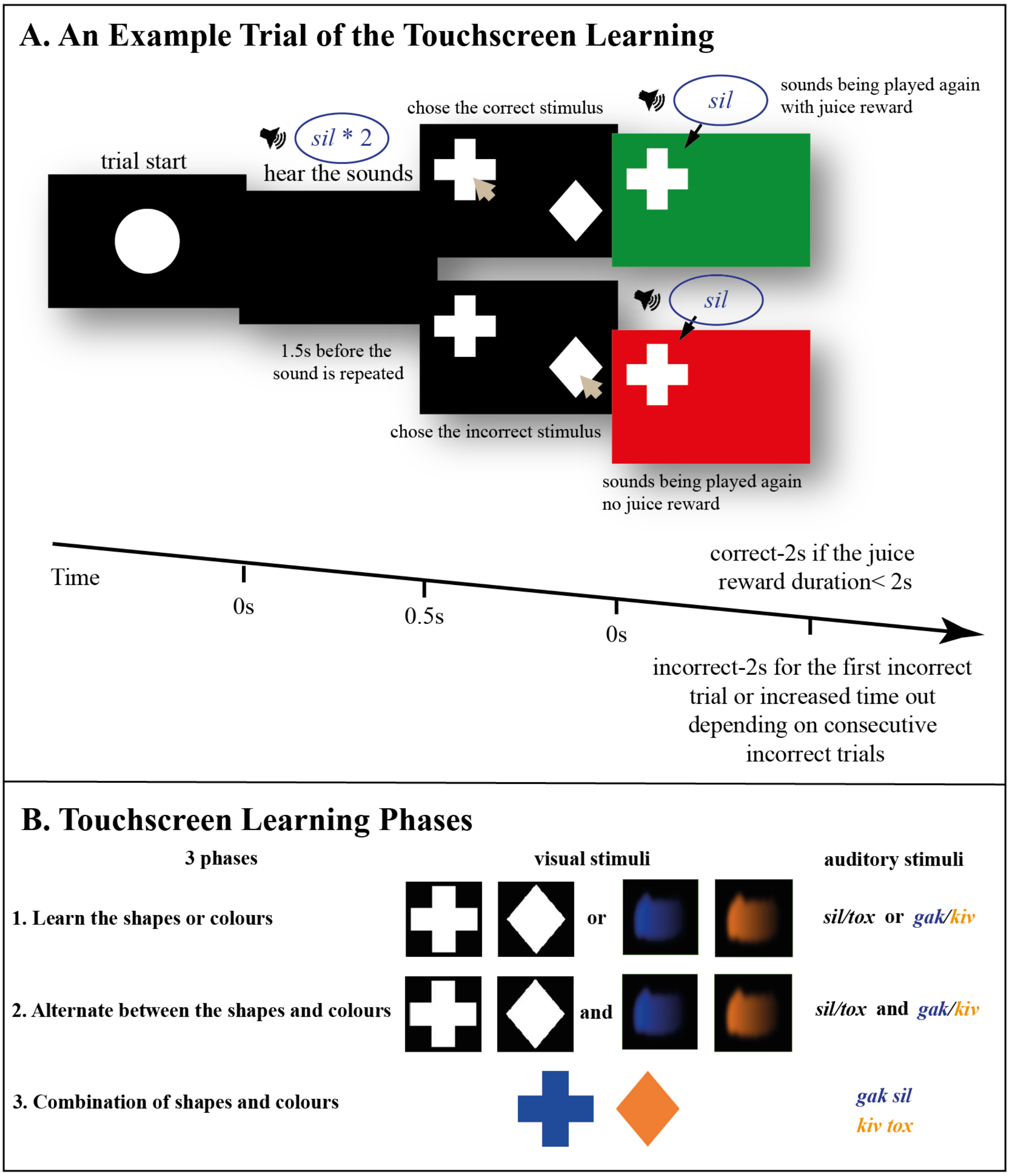
Following the third phase that the monkeys had to learn to combine the colour and shape features where sequences of two sounds identified objects by both colour and shape properties. Additionally, we attempted to introduce the correction trial on the touchscreen task. Specifically, if the monkey makes an incorrect choice in a trial, the trial is repeated one or more times with the visual stimuli presented in the same position on the screen. Furthermore, the sound referring to the visual stimulus is played the same as before, and the touchscreen task does not proceed to the next trial until the monkey makes a correct choice. We eventually set the correction trial to 1, meaning that if the monkey makes an incorrect trial, he would have only one chance to correct rather than keeping making the choice until the trial is correct just for getting the juice reward.

The monkeys could be moved to the next testing phase if they reached the threshold that have been set. However, more sessions were tested to ensure that the monkeys really learned the associations between the auditory and visual stimuli. The reasons that I often did not move the monkeys immediately to the next phase after they met criterion was for the following reasons. First of all, I want to make sure they really learned the associations between the audio/visual stimuli rather than showing the occasional results that they can be tested with next touchscreen learning phase. Additionally, I tested them with more sessions trying to maintain the good performance.

Regarding to the touchscreen paradigm, panel A in Figure 2.2. indicates the detailed process of the touchscreen task. The touchscreen was attached to the front door of the home-cage when we started the test. To start the trial, the monkey needed to push the white spot (used as a cue) in the centre of the screen, then the speech sound associated with the object (colour/shape) was played once/twice with an interval of 1.5 seconds in succession, followed by the stimuli (2 or 3 colours/shapes) being presented randomly on the screen right after the sound being played with no delay.

The interval of 1.5 seconds was only used when repeating the same sound twice during the colour or shape only learning. Notably, the delay of the same sound being played was used to help the monkeys to associate the sound with the visual stimulus rather than confusing the monkeys that the visual stimulus was associated with two sounds. When it came to the final testing phase which was the combinatorial learning phase, the interval of each sound was 0.75 seconds to ensure that the monkey could hear the sounds one after another immediately. Also, the intertrial interval was adjusted accordingly to the monkeys' behaviour as M1 was not affected by a long time out. However, M2 would leave the touch screen without engagement if there was a longer delay between each trial. That said, the more consistently incorrect trials that the monkey gave, the longer was the time out between each trial. By doing this, I was trying to encourage the monkeys to accomplish more correct trials. In addition, the stimuli were presented in different positions as the screen was divided into four quadrants or six areas equally as needed, meaning that the stimulus could be presented in any of the positions within these sections. By doing this, the monkeys could not rely on the spatial positions to pick the right stimulus and get the juice reward. If the monkey made a correct choice, the screen turned green, and the stimulus selected remained on the screen for as long as the time of juice reward was given. Additionally, the sound associated with the stimulus was played again to enhance the monkey's memory of the association between the audio-visual stimuli. Meanwhile, the juice reward was given to the monkey depending on the number of consecutive correct trials the monkey gave, the more consecutive correct trials, the higher amount of juice was rewarded. That said, the time of the juice reward was increased accordingly if the monkey gave consecutive correct trials. For example, the monkey would be given one second of the juice reward for the first correct trial, 2 seconds for the second consecutive correct trial, and three seconds or longer time for the third consecutive correct trial as the reward duration could be adjusted. In some cases, the monkey would be given more than 10 seconds if the number of consecutive trials was above 5. Afterwards, it moved to the next trial. The delay between each trial was 2 seconds if the monkey made consecutive correct trials after the juice reward was given.

However, if the monkey made an incorrect choice, the screen turned red with an interval of 2 seconds before it moved to the next trial, and the stimulus associated with the speech sound that was played before would be presented on the screen. By doing this, the monkeys were reminded of the correct association that would have led to reward. Afterwards, it moved to the next trial after the 2 seconds interval. As formerly mentioned, the delay between the trials varied depending on the consecutive incorrect trials the monkey made, meaning that the more incorrect consecutive trials, the longer delay it could be before proceeding to the next trial and the monkey was not rewarded for incorrect trials.



A. An Example Trial of the Touchscreen Learning

B. Touchscreen Learning Phases

Figure 2.2. Touchscreen task design. A) An example trial of the touchscreen learning. The procedure of an example trial is illustrated with the timeline which is presented in the line with an arrow. Notably, the time outs of correct and incorrect vary depending on the number of consecutive trials. B) Touchscreen learning phases. Three learning phases of the touchscreen task which are to learn the colours or the shapes, to alternate between the colour and shape versions and to learn the combination of the colour and shape referring to an object.

2.4 Discussion

This chapter presents the detailed methods and the task I developed for the touchscreen learning paradigm. Specifically, the monkeys were trained on the touchscreen task step by step, learning the task with gradually increased complexity.

Other studies such as the Medam study, their baboons needed to figure out the touchscreen learning by themselves without being trained step by step, and their baboons had unlimited time of access to the touchscreen (Medam & Fagot, 2016). However, the touchscreen paradigm I developed smoothly trained the monkeys step by step, although the time of the access to the touchscreen was limited considering of the animal welfare. Additionally, many cognitive studies on animals were conducted in zoos (or outdoor enclosure) or with characteristically designed 'testing box' and similar home-cage testing apparatus (Calapai, Cabrera-Moreno, Moser, & Jeschke, 2022; Egelkamp & Ross, 2019; Gazes, Lutz, Meyer, Hassett, & Hampton, 2019; Hopper, 2017; Nakamura et al., 2018). The varied experimental environments might have an impact on the animals' performance or the participation in the involved task. These were different from the experimental environment when I tested the monkeys on the touchscreen task as the monkeys involved in my study were tested in the home-cage in the colony, and the possibility that the monkeys being distracted might be increased by different factors such as people walking by, or the other monkeys being taken out from their home-cage. But I tried to minimise the effect of any possible distractions such as moving the monkey to the testing cage which was far away from other monkeys and using the audio-speakers to make sure that the sounds could be clearly heard by the monkeys.

In comparison to other studies, the significant difference of the touchscreen task I designed might be utilising the auditory and visual stimuli simultaneously and testing the animals with combined colour and shape features. Some other studies on animals and children used either visual or auditory stimuli only or tested the animals with less complex audio-visual associations or matching to goal experiments (Calapai et al., 2022; Ghasemian, Vardanjani, Sheibani, & Mansouri, 2021b; Medam & Fagot, 2016; Rey et al., 2012; Vonk & Rastogi, 2019). Although the complexity of the touchscreen task I designed might be difficult for the monkeys to perform, the monkeys were trained step by step with gradually increased complexity, leading to a smooth process of different testing phases.

Using this development, Chapter 3 examines the initial training performance with different methods, investigating each aspect which might cause an impact on the monkeys' performance. Chapter 4 examines whether the monkeys would show any bias to the shape or colour features by alternating the colour and shape only testing with different frequencies to enable the minimised possibility of the shape or colour bias. Finally, Chapter 5 examines whether the monkey could learn to identify an object by its shape and colour properties by hearing a short sequence of two sounds that refer to this object and whether the monkey could generalise the learning to novel combinations.

To conclude, this chapter lists the touchscreen setup and the task key components that the macaques participated in. In the next chapter, I describe the initial training results.

Chapter 3: Initial Training Design and Performance Assessment Methods

Abstract

To investigate the feasibility of the monkeys to do the initial task, we evaluated the characteristics of their performance on the touchscreen and implemented several methods to try to increase the monkeys' performance on the task. Both monkeys were able to perform hundreds of trials per session indicating a good motivation while performing the touchscreen task. Performance was variable throughout the testing session, but a permutation test method was able to identify 'good' performance sessions. We optimised the touchscreen training parameters for each monkey and simplified the touchscreen task from 6 sounds to learn to 4 sounds (referring to two colours and shapes).

3.1 Introduction

Many studies have demonstrated the success or feasibility of touchscreen training in macaques (Butler & Kennerley, 2019; Egelkamp & Ross, 2019; Medam & Fagot, 2016) and marmosets (Calapai et al., 2022; Nakamura et al., 2018). Not many of the studies are using auditory and visual stimuli simultaneously on the touchscreen task. Therefore, we might expect the monkeys to struggle with the task at least some of the time, highlighting ways in which the paradigm could be optimised for each monkey before proceeding to the latter stages of testing. Additionally, minor changes or differences in the paradigm that discourage the monkeys to complete more correct trials might also have some impact on the animals' behavioural strategy and the outcome of this research.

In this chapter, the initial touchscreen training results will be illustrated in more details. On top of that, I describe the approaches I have tried to improve the monkeys' performance on the touchscreen task.

3.2 Touchscreen Average Number of Trials Conducted per Session

The number of trials in different learning phases varies as the complexity of the touchscreen task was increased. In Phase 1, the monkey only needed to listen to a speech sound associated with either one of the colours or shapes. However, in Phase 2, the monkey needed to get used to the frequently switched versions of the colour only and shape only versions of the touchscreen task. In the final phase, the monkey needed to listen to two speech sounds referring to an integrated visual object with corresponding colour and shape properties. The complexity of the touchscreen task could possibly lead to the number of trials decreasing within a testing session. Next, I describe the initial testing results in more detail.

3.2.1 Phase 1 Initial Colour or Shape Training – Overview

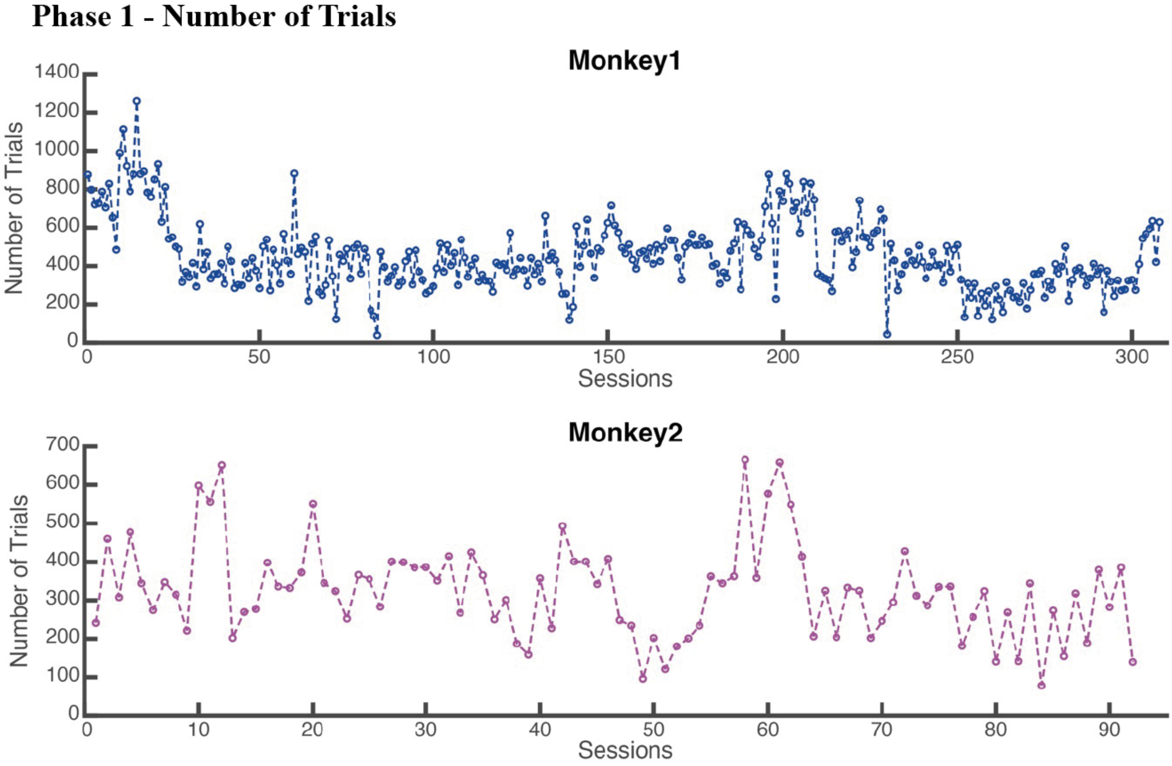


Figure 3.1. Summary of the number of trials from some of the sessions collected in Phase 1 including M1 and M2. The panel above in blue dotted line shows the number of trials over sessions in M1, and the panel below in pink dotted lines indicates the number of trials over sessions in M2. The number of trials varies in each individual session.

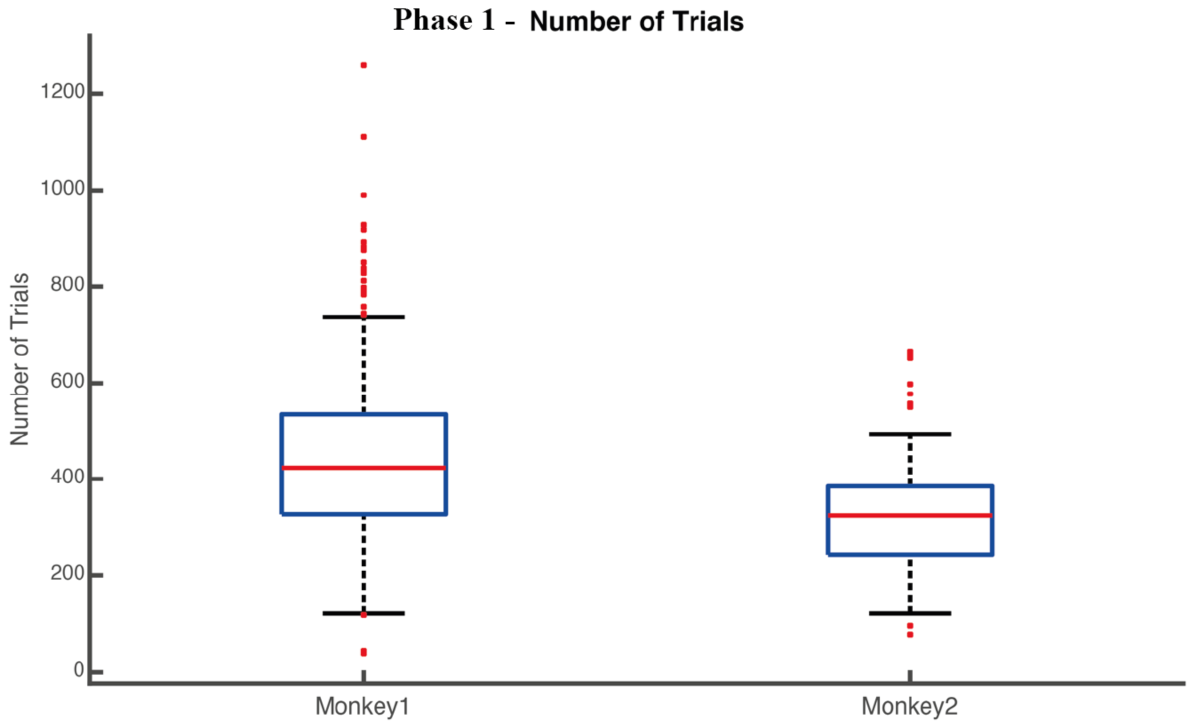


Figure 3.2. The box plot of the number of trials in both monkeys in Phase 1. M1 can do more trials within a session compared to M2, and the maximum number of trials within one session of M1 reaches to 1300 while M2 can do approximately 800 trials in maximum within a testing session. M1 does approximately 420 trials in average, and M2 does about 390 trials. There is not s significant difference between the number of trials among the two monkeys while being tested in Phase 1 (error bars are standard deviation).

As is shown in Figure 3.1. and Figure 3.2. above, the number of trials in the initial learning Phase 1 that both monkeys could perform were quite substantial in the 2-3 hours testing session. Both monkeys could perform hundreds of trials per session on average. Additionally, M1 did more trials on average compared to M2, and he could reach approximately 1300 trials within a session. This suggests that both monkeys were quite involved in the touchscreen learning task, and they could perform a large number of trials in each session. However, the initial training results also indicate the M2 on average performed fewer trials, and with this monkey I needed to ensure that the task was manageable so that it did not give up working on the touch screen. I also needed to optimise the reward and time out parameters, respectively, after a correct and incorrect trial so that M2 would not collapse in their performance. Therefore, both monkeys' performance needed to be optimised differently to ensure at least several hundred trials per testing session.

3.2.2 Phase 2 Colour and Shape Switching – Overview

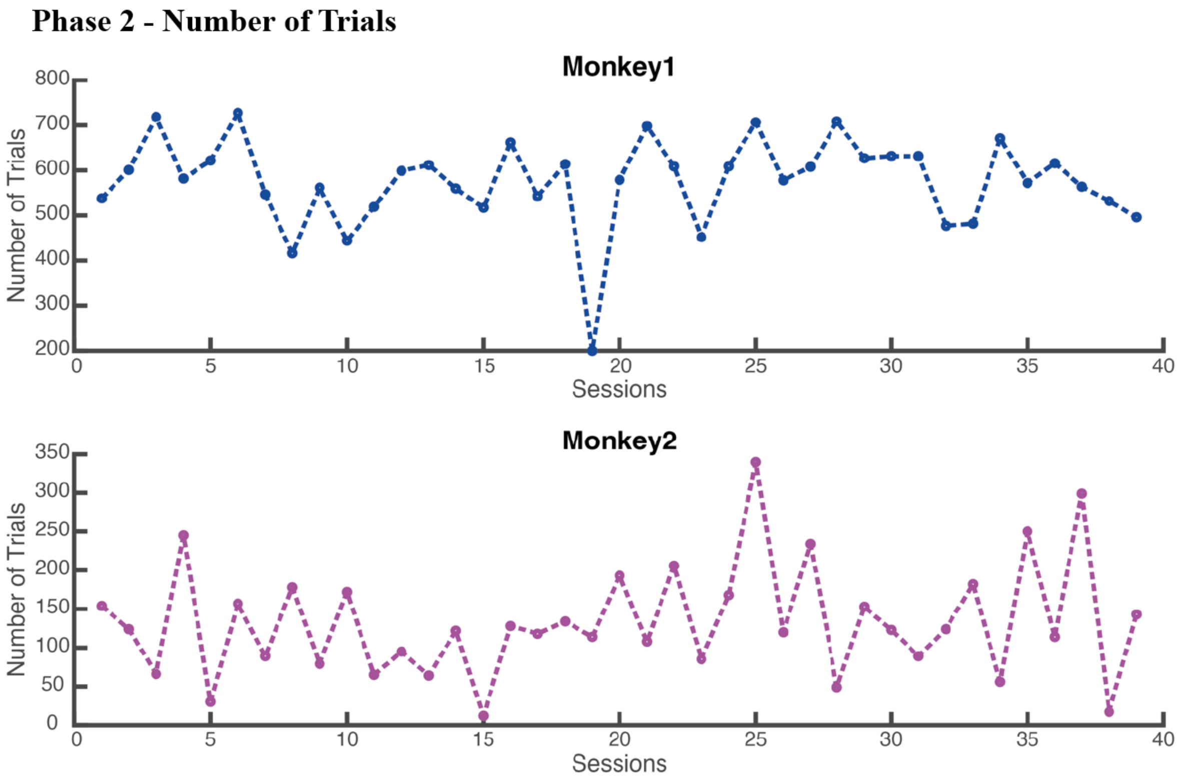


Figure 3.3. A summary of the number of trials from some of the sessions collected in Phase 2 including both monkeys. The panel above in blue dotted line shows the number of trials over sessions in Monkey 1, and the panel below in pink dotted lines indicates the number of trials over sessions in Monkey 2.

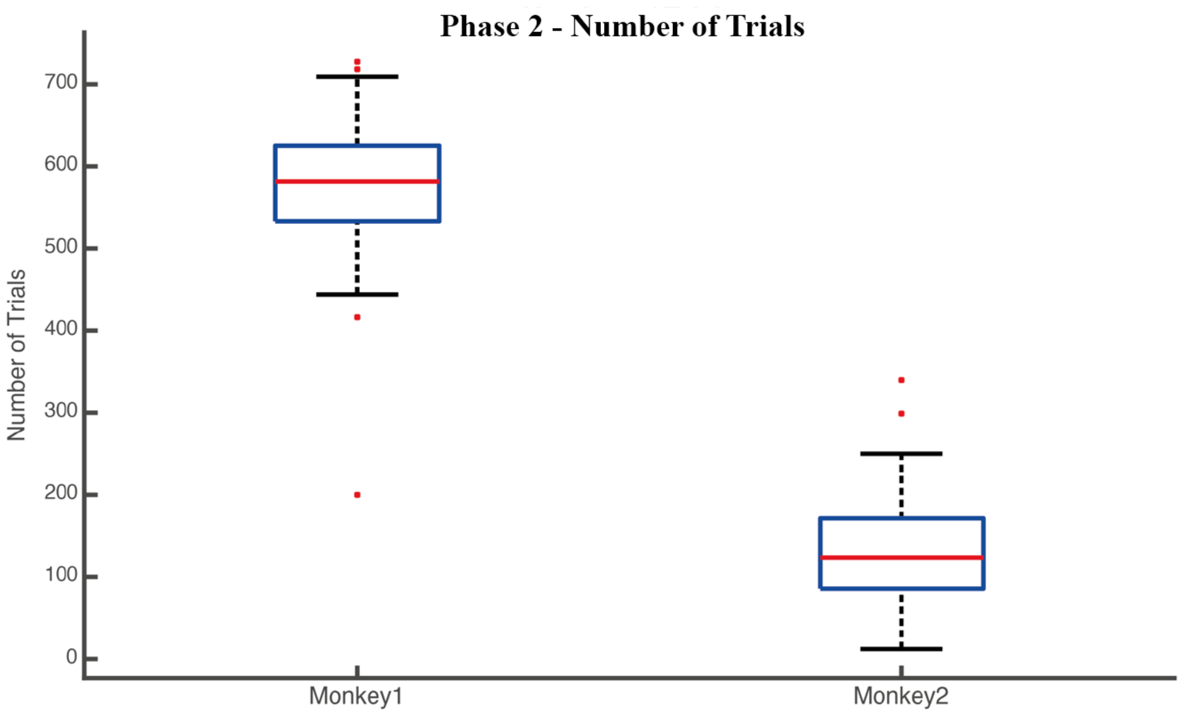


Figure 3.4. The box plot of the number of trials in both monkeys in Phase 2. Monkey 1 could do more trials within a session compared to Monkey 2, and the maximum number of trials within one session of Monkey 1 was more than 700 while Monkey 2 could do approximately 400 trials in maximum within a testing session. M1 did approximately 580 trials in average, and M2 did about 150 trials (error bars are standard deviation).

For the Phase 2 training, the task was switched between the colour only and shape only versions week by week, day by day or within the same day. As is shown in the figure above, there is a significant difference in the number of trials among the two monkeys when being tested in Phase 2. M1 stayed stable in terms of the number of trials from Phase 1 to Phase 2 (increasing slightly to an average of about 600 trials). By comparison, M2 dropped down the number of trials produced for Phase 2 which was approximately half the number of trials compared to Phase 1 on average about 100 trials. This might be explained by the increased difficulty of the touchscreen task itself as the colour and shape versions were switched frequently during the testing days, which could have made some impact on the monkeys' behaviour, leading to the decreased number of trials per session for M2 but not M1.

3.2.3 Phase 3 Combinatorial Learning - Overview

The data of Phase 3 was only collected from M1 before this monkey needed to complete all testing. Additionally, there was no data collected from M2 on Phase 3 due to some unexpected issues which led to the monkey completing all procedures early in agreement with the Home Office.

Figure 3.5. shows the number of trials collected in M1 during Phase 3. There were 49 sessions in total including some of the sessions tested with familiar and novel conditions as described in more detail in Chapter 5. Given that this may have been the most difficult part of the behavioural experiments, the combinatorial learning testing phase, the number of trials in each session in M1 were fewer compared to Phase 1 and 2.

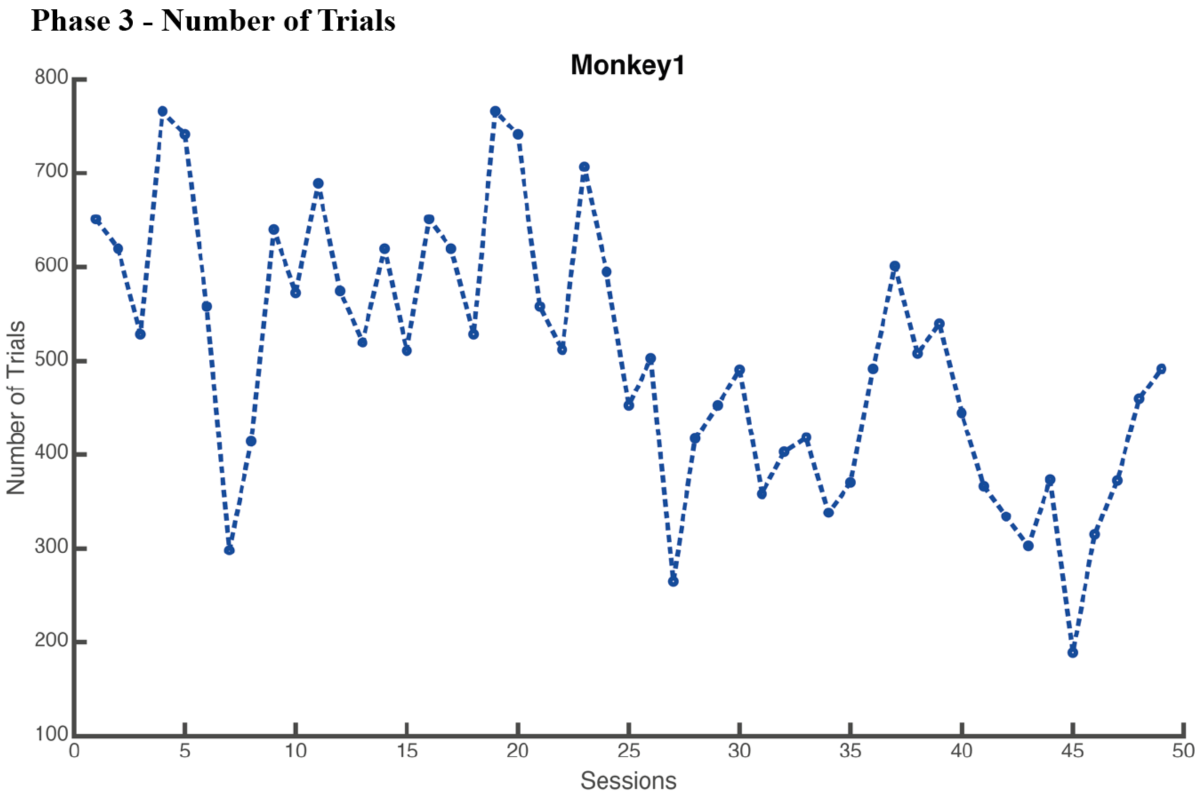

Phase 3 - Number of Trials

Figure 3.5. A summary of the number of trials collected in Phase 3 in Monkey 1.

The figure below presents a box plot of the number of trials of M1 in Phase 3. The monkey stayed stable in terms of the number of trials, which might suggest that the monkey was still quite motivated while performing the touchscreen task even the difficulty was increased with regards to the combined auditory and visual stimuli.

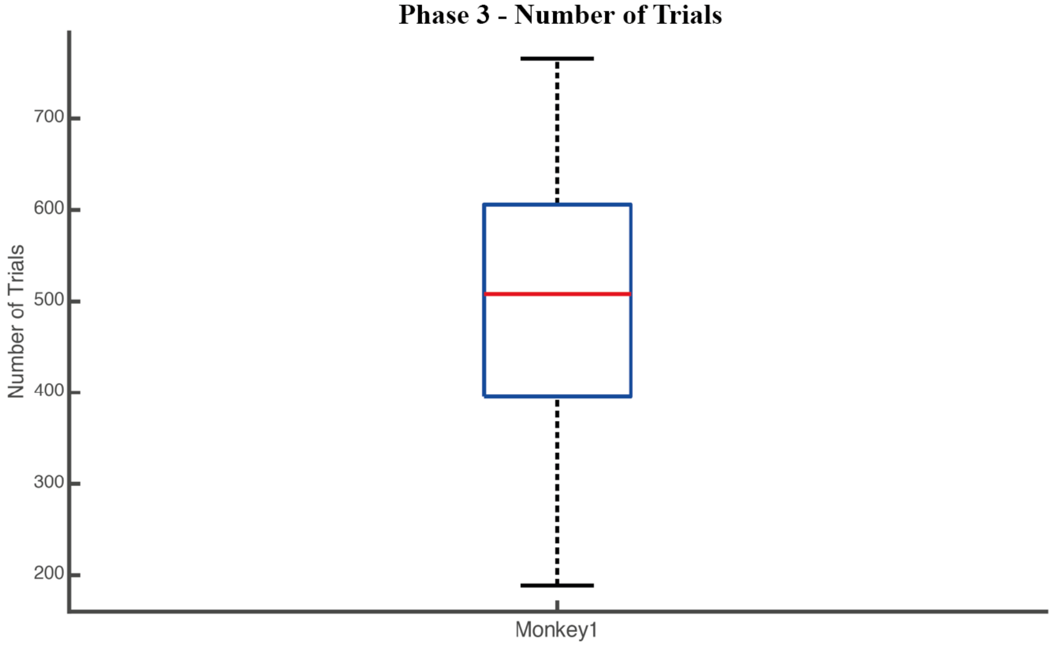


Figure 3.6. The box plot of the number of trials in Monkey 1 in Phase 3 (error bars are standard deviation).

3.3 Evaluating Performance Session by Session – Sliding Window Analysis

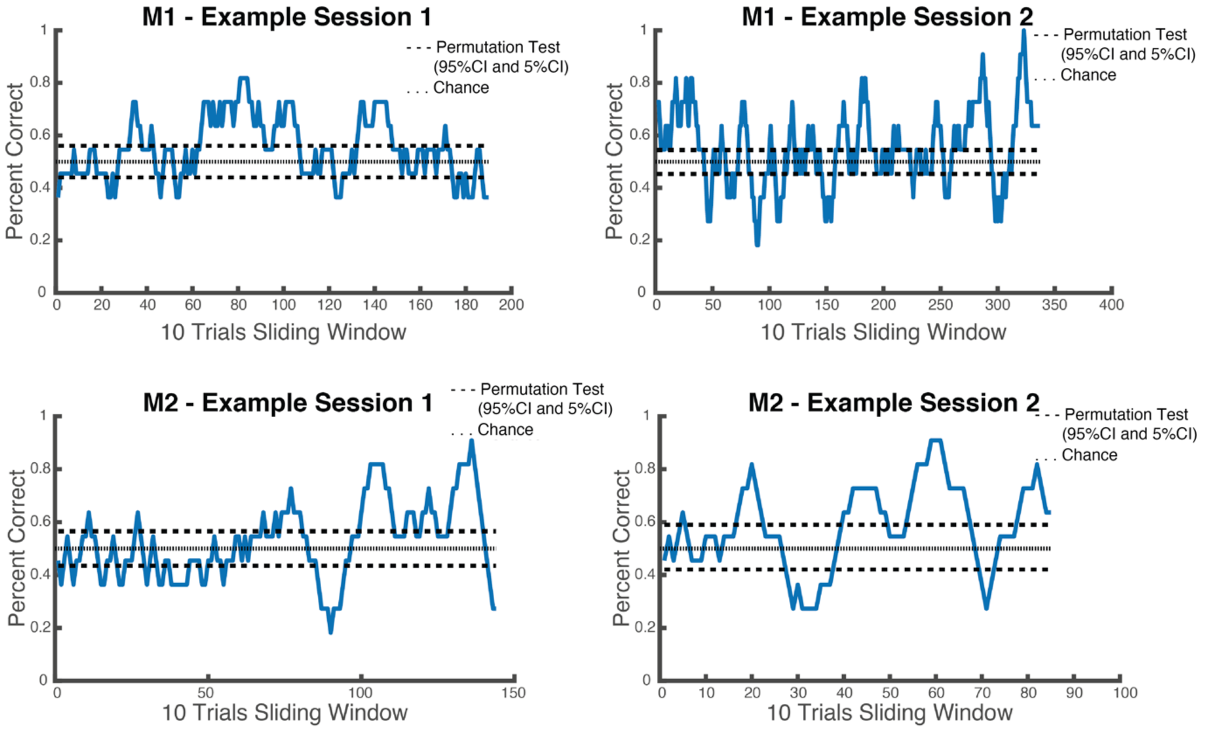


Figure 3.7. Example testing session in both monkeys. Sliding through each 10 trials, shifting by 1 trial each time in each session. And we applied a permutation test shuffling the monkey's own response for 10 thousand times within the testing session to give a null distribution of

performance that identifies the performance confidence limits as represented for 95% CI and 5% CI. The dash lines on top and below indicate the chance levels given the permutation test which calculates the chance levels relative to confidence intervals (95% CI and 5% CI). The dotted line in the middle indicates the chance performance which is 50% as the stimuli are 2 in these testing sessions. The blue line in each panel presents the percent correct. The x axis shows the number of trials over time, and the y axis shows the percent correct of the number of trials based on a 10-trial sliding window.

As is shown in the figure above (Figure 3.7.), both monkeys performed above chance the majority of time within a testing session, however, performance was quite variable within a session with periods of chance performance alternating with periods of very good performance. For instance, using a 10-trial sliding window analysis to evaluate percent correct during each session, there are often bursts of performance reaching to 90 percent correct at some time periods. This made it difficult to rely on just the mean percent correct performance throughout each session or the max percent correct performance. The mean percent correct performance was often close to chance levels (50 % correct) or the max performance was close to ceiling (> 85% correct).

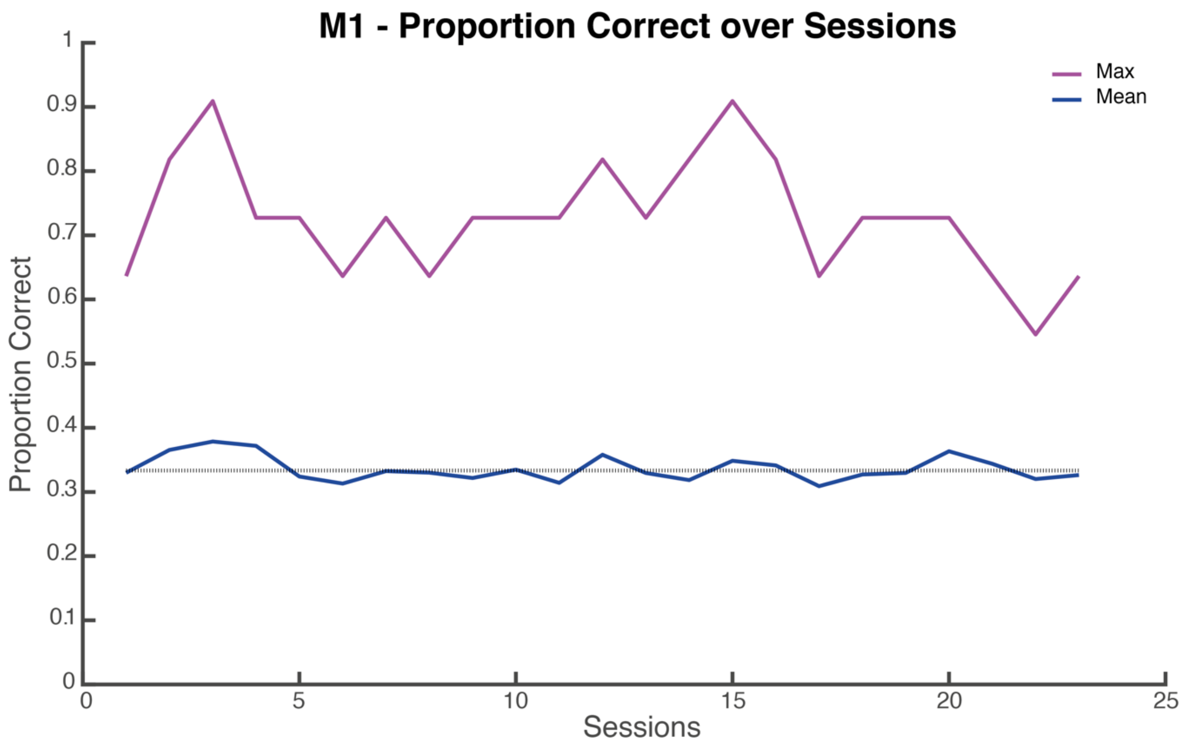


Figure 3.8. Proportion correct over sessions in M1. The blue line indicates the mean performance in each of the session, and the red line shows the maximum percent correct based

on a 10-trial sliding window analysis, calculating the performance of 10 trials in a row within that session. The dotted line presents the chance level which is 33.3% as the stimuli were 3 among these testing sessions.

As is shown in the figure (Figure 3.8.), the mean performance across sessions was close to chance (33% when there were 3 sounds; 50% when there were 2 sounds). Neither measure was informative enough to tell whether the monkey performed well across these sessions as the mean performance is fluctuating up and down. Therefore, to get more information with further analysis of the data, we conducted a permutation test that allowed us to identify the 5% and 95% confidence intervals (horizontal dashed lines in the example sessions in Figure 3.7.) as follows.

Specifically, the permutation analysis shuffles the monkey's own responses within its own session 10 thousand times to give a null distribution of performance that identifies the 'good' performance confidence limits. In essence, the permutation test simulates a monkey that does not know what the correct conditions are but uses the monkey's own responses and the stimulus conditions that were experienced during the testing session.

For each session, a set of random responses is generated 10 thousand times, giving a null distribution of random responses, and from this, the 5% and 95% confidence interval can be calculated session by session. Afterwards, the number of times that the actual performance, by which was based on the sliding window analysis, crossed either the 5% or the 95% confidence interval was counted. Similarly, the area of these breaches crossed the 5% and 95% confidence intervals was also used as a metric.

This allowed us to identify the trials for which the monkey performed above chance. However, because a false breach above chance might be possible, we compared the number of breaches both in terms of area and number of trials above chance to the below chance breaches. By doing this, if the monkey was truly guessing, we should see as many breaches below the 5% confidence interval as above the 95% confidence interval. However, we reasoned that if the monkey performs well above chance, performance based on the sliding window analysis should go above the 95% CI more often than it drops below the 5% CI. As is shown in the figures (Figure 3.9. and Figure 3.10.), the number of breaches and the area of breaches above 95% CI or below 5% CI was calculated. Additionally, for the sessions identified as 'good' performance sessions, the number or area of breaches above 95% CI in majority of the sessions was higher than the ones below 5% CI. I also conducted another permutation test (a binary coin flip model) and the results were similar in identifying sessions with 'good' performance. Because the first method relies on the monkey's actual responses and the stimulus conditions experienced, I opted to use the first approach (permuted confidence intervals) to identify good performance sessions for further analysis and to progress the monkey to the next testing phase, e.g., when performance identified as 'good' was identified for the majority of the testing sessions in at least 2 weeks of testing.

I tried to test the monkeys with more sessions to encourage the monkeys to maintain stable and 'good' performance by our significance criteria across individual sessions. Thus, I did not immediately move the monkeys on when they initially reached out the threshold (6 out of 10 good sessions). As formerly mentioned, the reasons are to make sure that they learn the associations, and to maintain their performance.

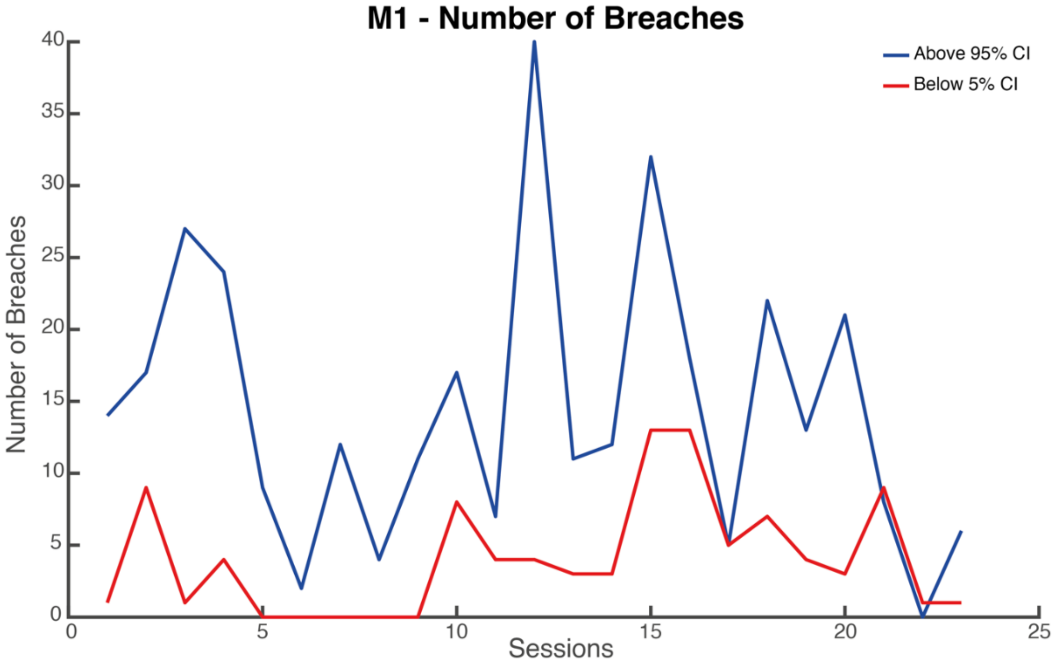


Figure 3.9. Number of breaches in M1. The blue line indicates the number of breaches above 95% CI and the red line presents the number of breaches below 5% CI (Example good testing sessions in Phase 1).

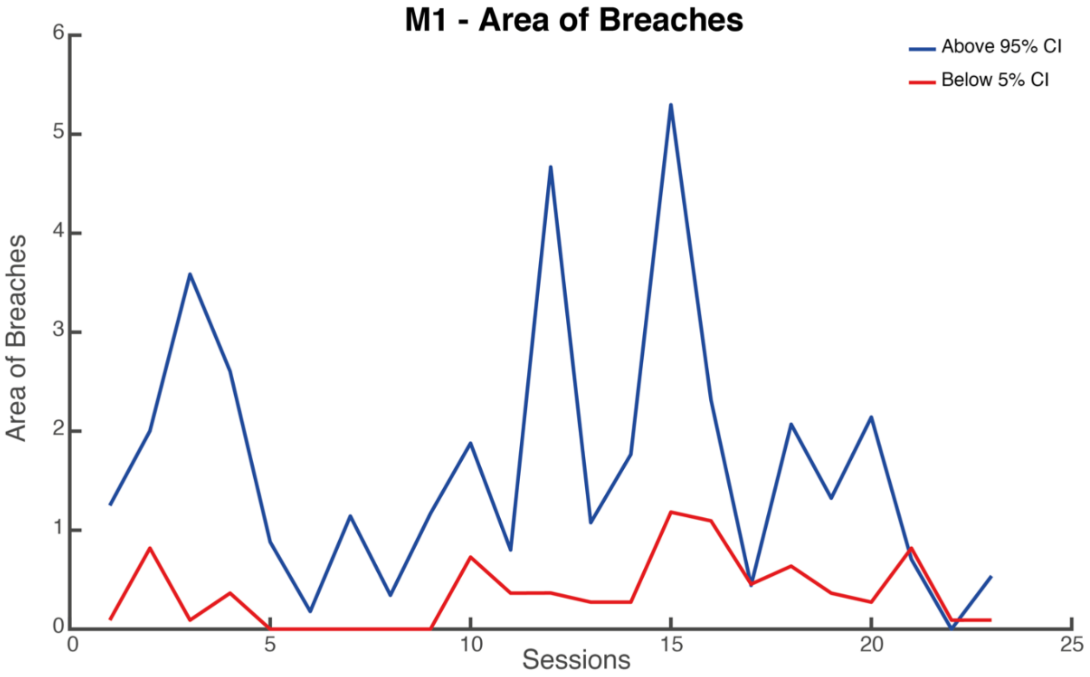


Figure 3.10. Area of breaches in M1. The blue line indicates the area of breaches above 95% CI, and the red line presents the area of breaches below 5% CI (Example good testing sessions in Phase 1).

Figures (Figure 3.9. and Figure 3.10.) show M1's performance across sessions during Phase 1. Summarised as the number of trials breached above or below the confidence intervals. The blue line is the number of trials above chance, and the red line is the number of trials below chance.

Figure 3.11. shows the difference in the number of breaches between the above and below chance. Whenever the breach above chance is higher, the sessions are marked in red stars indicating significant above chance performance. There are some sessions showing the difference that is below 0, this is because the number of breaches below 5% CI was higher than the ones above 95% CI, indicating that the monkey did not perform well within that session as there were many time periods that the monkey's performance drops below the chance or the above versus below chance performance do not differ. I used a threshold of above breaches of at least 10 to identify good sessions.

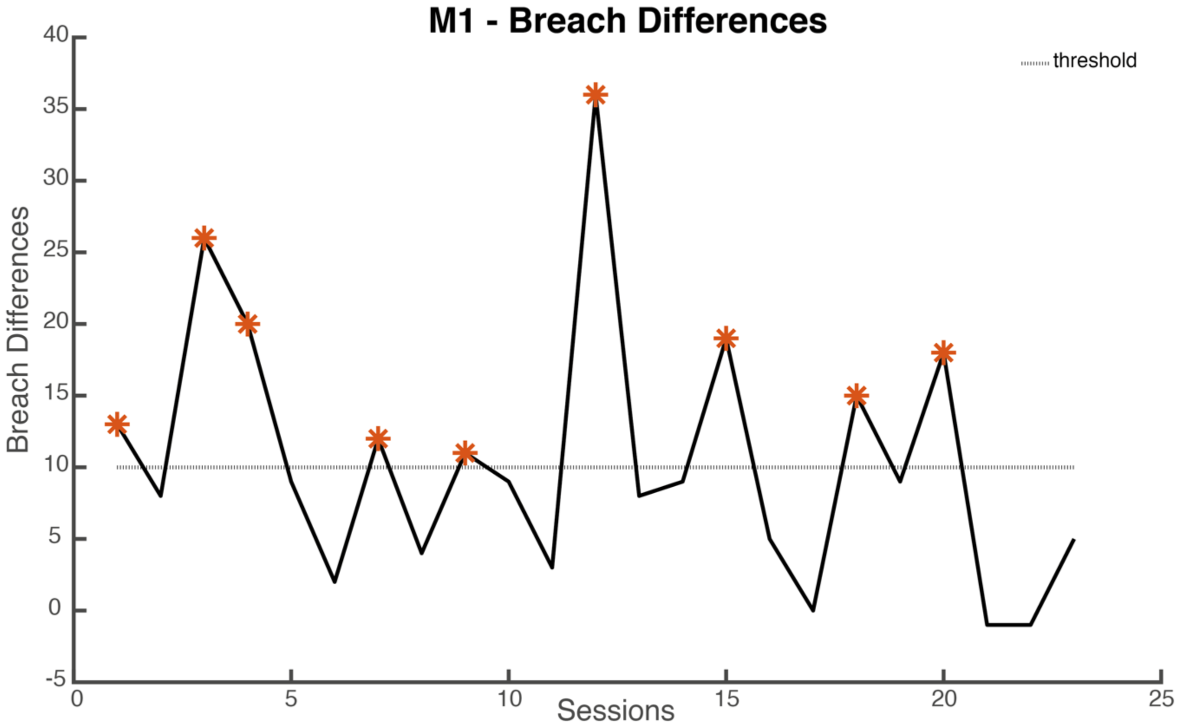


Figure 3.11. Difference in number of breaches in M1. The threshold is 10 presented in dotted line, and the stars marked in red indicate significant sessions that the monkey reached the threshold. The difference in number of breaches was calculated with the number of breaches above 95% CI minus the number of breaches below 5% CI.

3.4 Evaluating Monkey 1's Performance with Foil Visual Stimuli - Distractors

In the initial stages of training, we attempted to identify whether the monkeys could correctly identify the visual objects associated with the 3 sounds, by implementing several visual foil stimuli never associated with the sounds. Here the extra visual stimuli were used as distractors. We first tried to just introduce the 3 visual distractors to the monkey, and no sounds were associated with any of the distractors. In the meantime, the 3 visual distractors were presented on the screen with the stimuli, meaning that 6 shapes were presented on the screen at the same time. However, only 3 of them were the stimuli, and the monkey needed to select from the 3 visual stimuli with juice reward given if the monkey performed correctly, while ignoring the 3 distractors as they did not represent anything meaningful for reward but were there just to distract the monkeys or identify if the monkey was not paying attention to the task. We found that the monkey quickly learned to ignore the visual foils. Thus, we tried to introduce another 3 sounds associated with the 3 distractors for some sessions to assess whether the monkey could also learn to ignore them. Notably, the 3 visual distractors were associated with 3 sounds and were used with the ‘visual stimuli’. However, the visual stimuli were switched to the ‘distractors’ without any sounds being played when they were presented on the screen. By doing this, the character of the visual stimuli and the distractors were alternated. We also found that the monkey learned not to select among the ‘visual stimuli’ in a few testing sessions. This suggests that the monkey could distinguish the sounds being played to the ones referred to the true visual stimuli that they were associated with and which would lead to reward if chosen correctly. Therefore, the distractors were only used when necessary, but ultimately did not seem to be useful given that the monkey could easily learn to ignore them.

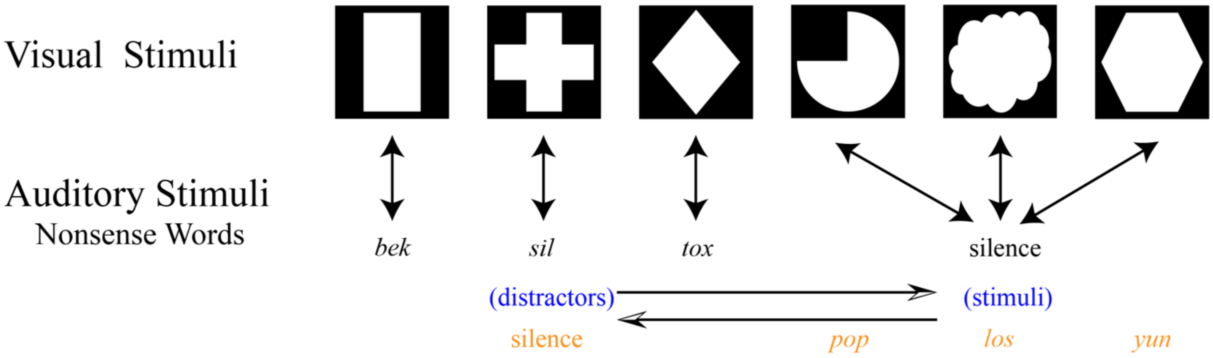


Figure 3.12. Using additional foils: example approach to improve the performance. The visual stimuli on the top of the figure are associated with the auditory ones at the bottom. Three visual stimuli were presented on the screen at the same time in different positions. The first 3 (rectangle, cross and diamond) on the left side and the other threes (partial circle, cloud and hexagon) on the right side were grouped into two groups and were used as stimuli or distractors.

Figure 3.12. illustrates an example approach that we tried to assess whether the monkey could ignore the visual foils. We firstly tried to present all the visual stimuli on the screen at the same time to see if there is any difference compared to only three visual stimuli being presented on the screen. In the meantime, the additional three stimuli used as distractors were not associated with any of the sounds. And we found that the monkeys quickly learned to ignore the distractors just within few testing sessions, meaning that they could identify the sounds being played were corresponded to specific visual stimuli, rather than picking randomly among these stimuli. Following by switching between the visual stimuli and the distractors, we tried to use the visual stimuli (rectangle, cross and diamond) as distractors and the distractors (partial circle, cloud, and hexagon) as the stimuli.

The result below shows an example session of the number of visual stimuli selected in Monkey 1 tested with 6 shapes with the total number of 175 sessions.

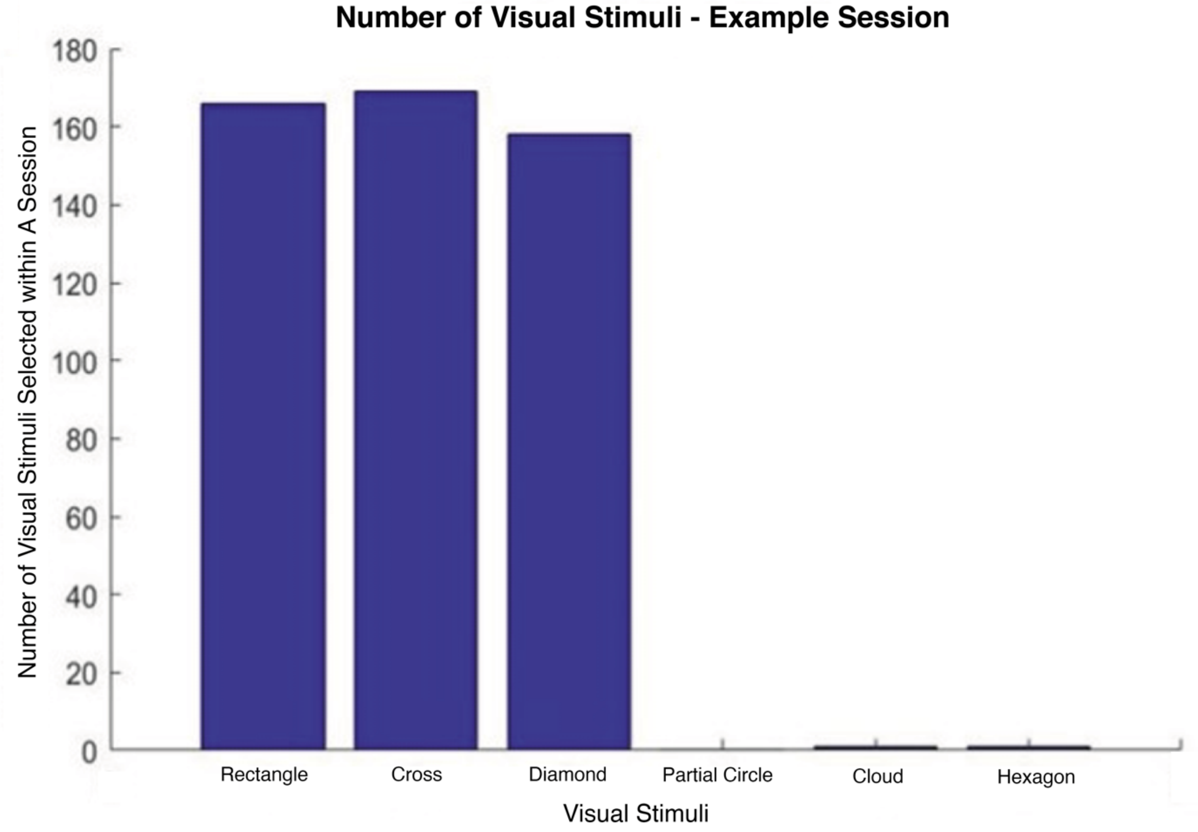


Figure 3.13. Example session of the number of visual stimuli selected. 6 shapes in total, and 3 of them are visual stimuli (rectangle, cross, and diamond) and another 3 are distractors used as foil stimuli (partial circle, cloud, and hexagon).

As is shown in the figure (Figure 3.13.), the first 3 shapes were used as the stimuli, and the extra 3 shapes were the distractors. Additionally, the extra shapes implemented into the touchscreen task were merely or not selected by the monkeys as they are used as distractors without giving the juice reward. As mentioned before, the distractors were only used when needed as the monkey quickly learned to ignore the distractors. This also indicates that the monkey can learn the associations between the auditory and visual stimuli.

3.5 Efforts to Improve Performance: Optimising Reward and Time Outs

I aimed to optimise the performance for each monkey individually. For instance, I firstly aimed to increase the juice reward for greater consecutive correct trials. I also increased the time outs to reduce the number of incorrect trials. M1 was not bothered by a longer time out even it was 3 or 5 minutes. However, M2 was more fragile and could be affected by the longer time out as it would not be interested in performing the touchscreen task if there was a long delay between the trials. In the initial training sessions, M2 would often quit the testing sessions if the time out was too long. Furthermore, our computer program monitored the performance online to give the monkey a longer penalty time out when performance falls to chance. Namely, a sliding window analysis of 5 trials was used during the testing session to assess the performance, and the screen would be shut down for 60 seconds or more if the monkey dips below 60% correct performance. By doing this, we aim to improve the monkeys' performance and encourage them to do more correct trials on the task when they are being tested.

The figure (Figure 3.14.) below shows the performance for breach trials tested in M1 with 2 visual stimuli, and the reward as well as the time outs were optimised. Specifically, the juice reward was increased accordingly to the consecutive correct trials that the monkey made, meaning that the more consecutive correct trials, the more juice reward was given to the monkey. Additionally, the time outs were increased by the number of incorrect trials. If the monkey consecutively made incorrect choices, the delay between trials was increased and it could be maximumly increased to 5 minutes, meaning that the screen was shut down for such a long period to avoid too many incorrect trials. The performance was increased although it was not significant, indicating that optimising the juice reward and the time outs did not help to improve the monkey's performance considerably.

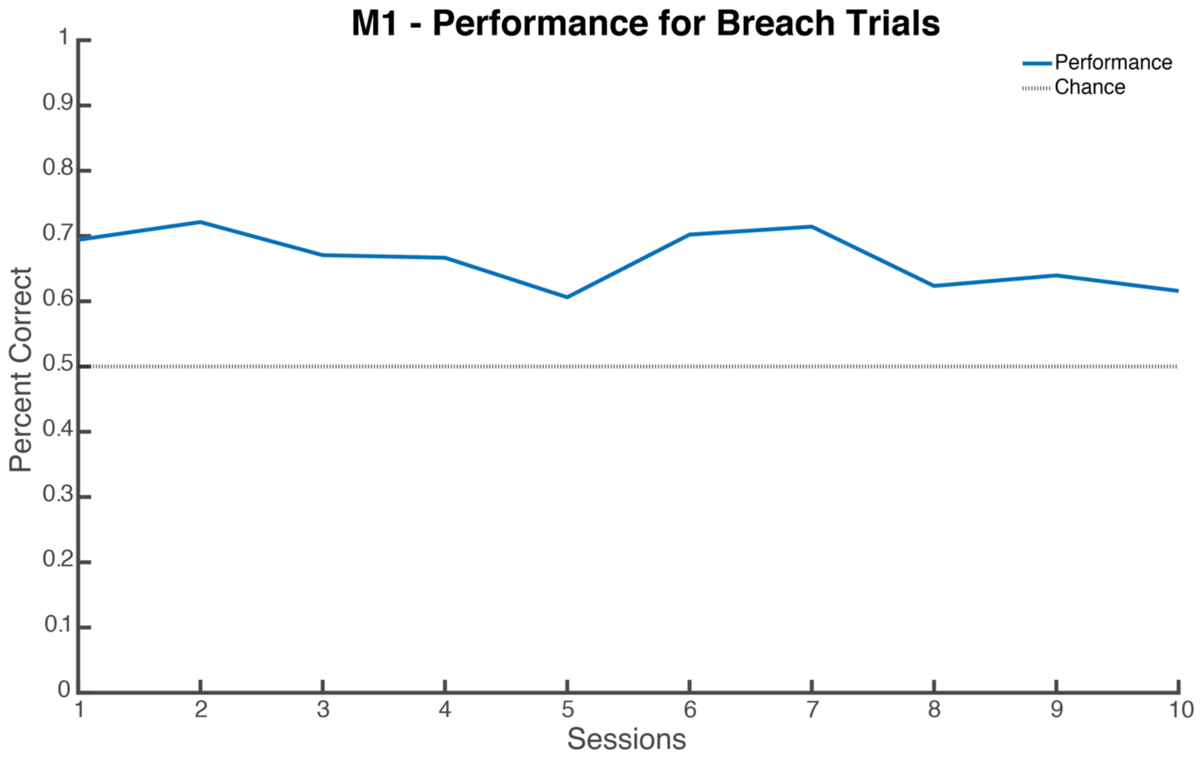


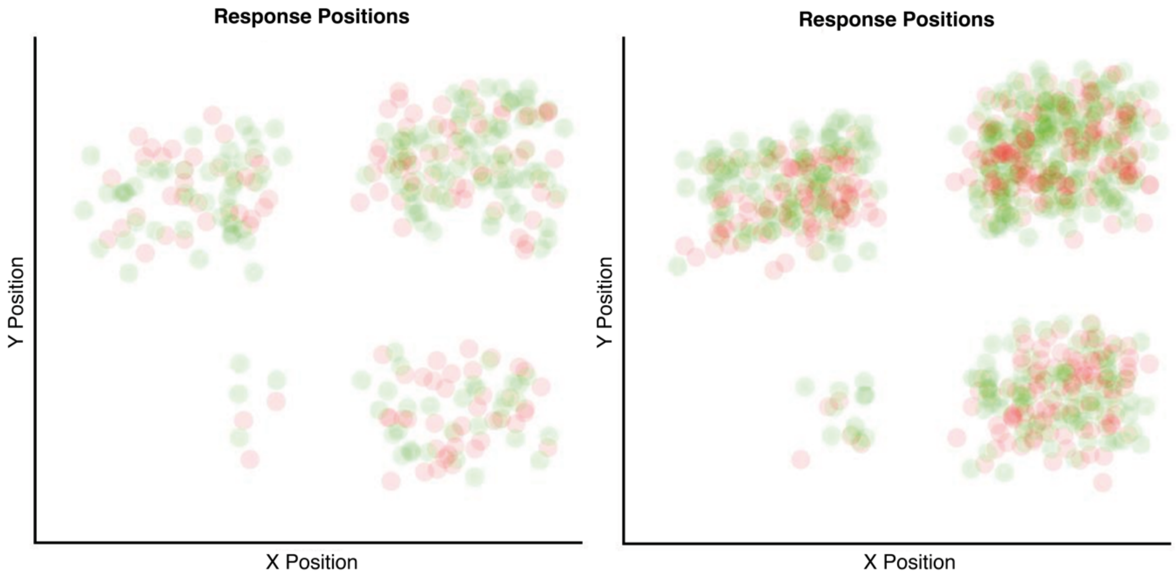
Figure 3.14. M1 – Performance for Breach Trials. Sessions tested with 2 visual stimuli, and the chance level here is 50%. The blue line shows the performance, and the dotted line indicates the chance level based on the stimuli presented of the testing session.

3.6 Efforts to Adjust for the Monkeys' Preferred Touching Positions

I also investigated the response positions that the monkeys gave during the testing sessions. Specifically, the screen is divided into 4 equal quadrants. And the visual stimuli are presented on the screen randomly even within the same quadrant. As is shown in the figure (Figure 3.15.), the red and green solid dots indicate the correct and incorrect responses, respectively. Both monkeys seemed to have a preference on the right side of the screen as more response were given on the right side. Additionally, both monkeys seemed to be right-handed as they were using their right hands while performing the touchscreen task. However, there seems to be a difference between the two monkeys as M1 does not respond to the left bottom as much as the other quadrants, and there are not many responses on the top left of the screen in M2. This might be because of the preference and the ways of how they manage to touch the screen. Another possibility could be the nature of the smaller size of M2 that it might not be that easy for M2 to reach the top left of the screen.

Additionally, if the monkey has a preference of the positions that he would like to select among the 4 quadrants, the responses could be affected no matter if it is correct or not. For example, if the stimulus is presented on the left bottom of the screen, the monkey might ignore where this visual stimulus is presented and touch any other positions of the screen to complete the trial. In this case, the trial would be an incorrect trial as the monkey does not pick the correct stimulus associated with the sound that has been played. After the monkey makes the response, the touchscreen task will continue to the next trial with a short time out, if this is the first incorrect trial that the monkey makes. This might be explained by the behavioural strategy that the monkey tries to get the juice reward as soon as possible while ignoring the stimuli in the positions he does not prefer. Moreover, the straight bars of the home-cage may have caused some impact to the monkeys' behaviour as the restrictions of these straight bars may lead to the limited access to the touchscreen during their testing sessions.

M1 - Response positions



M2 - Response positions

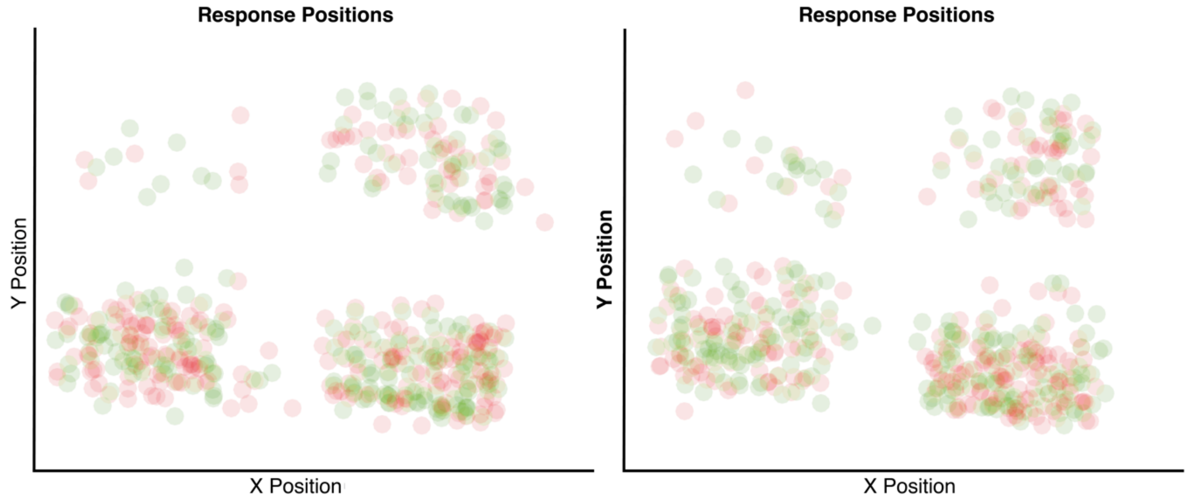


Figure 3.15. Example sessions of response positions in M1 and M2. The top two sessions show the response positions of the touchscreen learning in M1 and the panel below show the response positions in M2. The screen is divided into 4 quadrants equally according to the pixels of the screen (top left, top right, left bottom, right bottom). The green and red solid circles indicate the correct and incorrect responses, respectively. Additionally, the x and y position in the axis are scaled according to the pixels of the touchscreen that has been used for testing.

After figuring out the monkeys' response positions on the touchscreen, I first tried to not present stimuli in the quadrants that the monkeys give fewer response to encourage them to do more correct trials. For example, the top left quadrant was blocked meaning that no visual stimuli would be presented in that quadrant during the testing. However, this did not help with the monkey's performance. Thus, the following step was blocking the quadrants that the monkey had a preference to. For example, blocking the top right or the bottom right quadrants of the screen to make the monkey touch the left side of the screen. By doing this, the monkey needed to be concentrated on the task while being tested, and they had to touch the screen to the left side of the screen to make a correct trial with the juice reward. By doing this, the preference of the quadrant might be balanced. However, this did not help the monkey perform better and the monkey performed at chance level in most of the sessions. Thus, although testing accommodating the touch areas of the monkey or encouraging it to touch areas that it would tend not to, I ultimately continued with Phase 1, 2 and 3 training without blocking visual stimuli from being presented in specific parts of the touch screen quadrants. This may have led to more incorrect trials, for instance, the monkey got a trial incorrect because it was less comfortable for the monkey to touch the lower left quadrant furthest away from their hand.

Figure 3.16. presents the performance for breach trials in M1 with 2 stimuli (2 shapes or 2 colours) when the top left quadrant was blocked, meaning that the visual stimuli were not presented in this quadrant during the testing. The performance was neither increased nor decreased but stayed stable around chance performance. This suggests that blocking the quadrant that the monkey gave fewer responses did not help in relation to improve the monkey's performance.

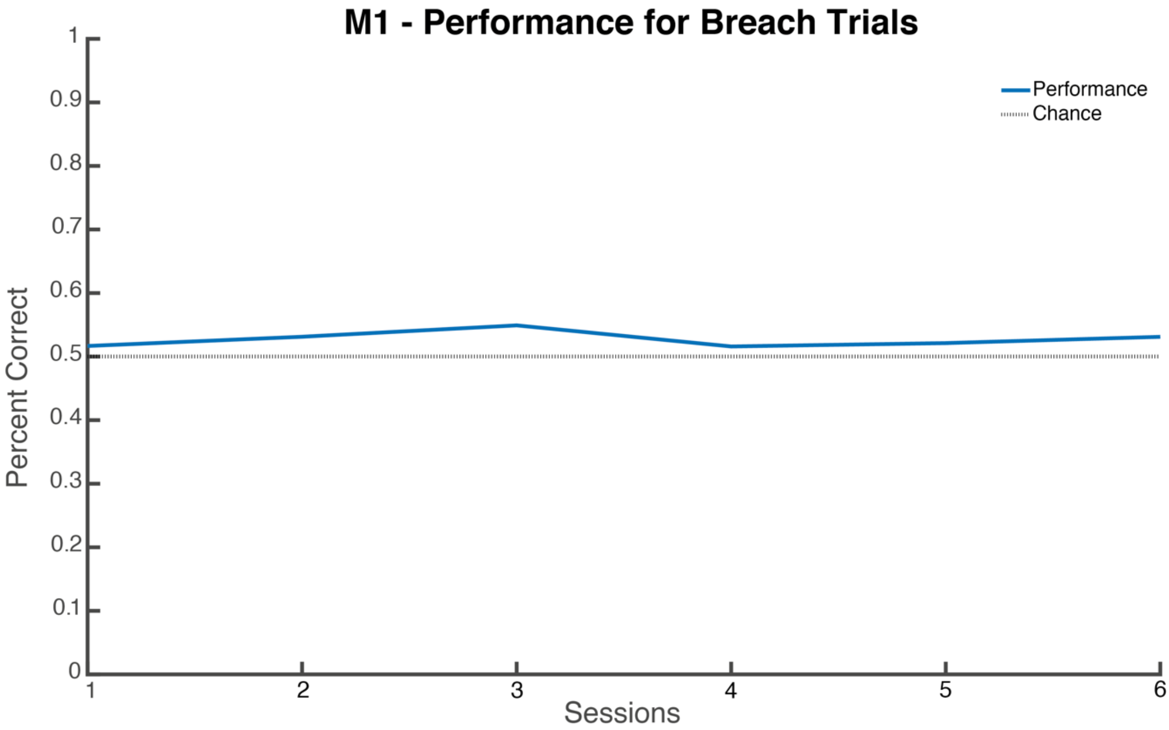


Figure 3.16. M1 – Performance for Breach Trials. Sessions tested with the top left quadrant blocked. 2 visual stimuli are used in these sessions and the chance level is 50%. The blue line shows the performance, and the dotted line indicates the chance level based on the stimuli presented of the testing session. Note that the monkey performed poorly when it did not have access to the touch areas that it preferred to touch.

The figure (Figure 3.17.) below shows the performance of each stimulus with the top left quadrant blocked when M1 was presented with 2 shapes or 2 colours. As is shown in the figure, there are no significant differences between the percent correct of each stimulus (paired t-test, t(5) = 0.8222, p = 0.4484). The percent correct of the stimulus blue is slightly higher than the orange. Similarly, the monkey performed slightly better on the diamond rather than the cross. The percent correct of each stimulus does not vary much when the top left quadrant was blocked, and the monkey's performance was not affected by the changes.

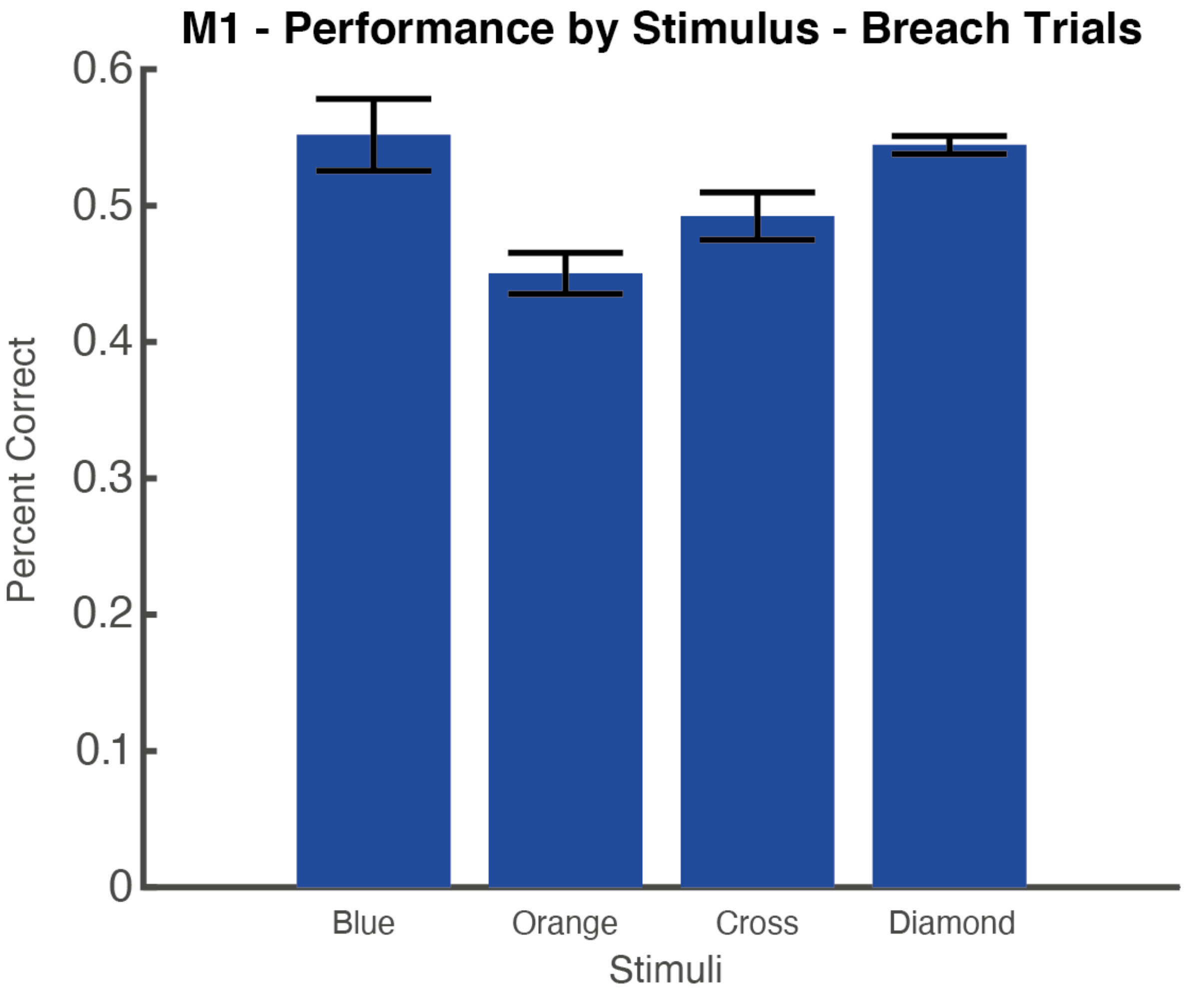


Figure 3.17. M1 – Performance by Stimulus – Breach Trials. Sessions tested with either 2 shapes or 2 colours. Blue and orange are the colour stimuli, and cross and diamond are the shape stimuli (paired t-test, N = 6, t(5) = 0.8222, p = 0.4484, error bars are standard deviation).

Figure 3.17. shows first indication that the monkeys might be showing a bias towards some of the colours or shapes, which I will study more systematically and try to balance before moving the monkey to the final Phase 3. I also investigated the reaction time that the monkey responded to each of the stimuli. Figure 3.18. shows the reaction time of each stimulus in seconds. These sessions were conducted with the quadrant on the top left blocked, and either 2 shapes only or 2 colours only were presented on the screen during the testing. As is shown in the figure, there are not many differences between the reaction time of each stimulus, the monkey responded to either the shapes or colours on average within 1 second, meaning that the monkey did not show significant difference when made response to the shapes or colours.

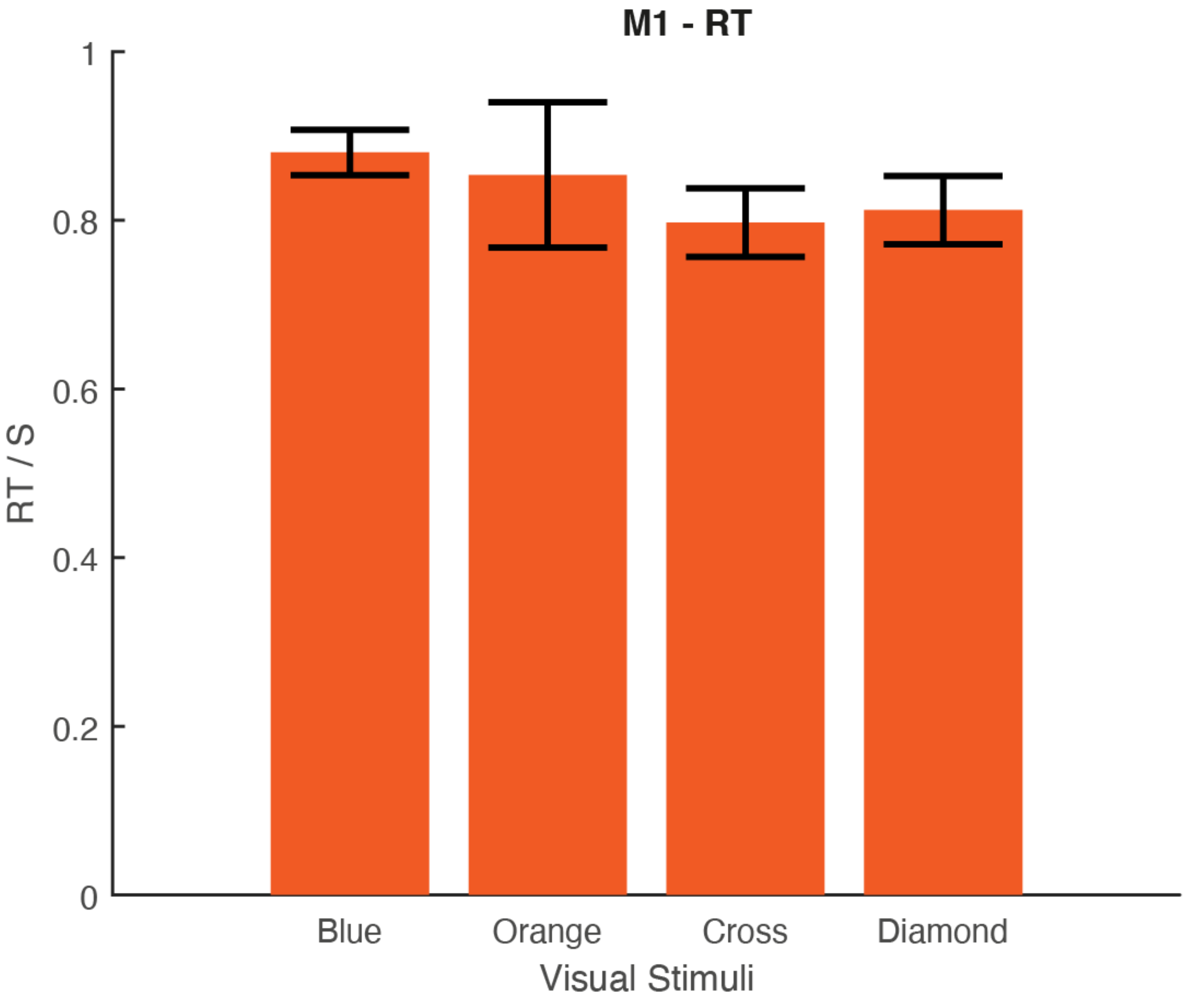


Figure 3.18. M1 – Reaction Time of Stimuli. These sessions were tested with either 2 colours only or 2 shapes only. The reaction time is shown in second (error bars are standard deviation).

3.7 Efforts to Improve Performance: Testing in Different Time Periods

Another attempt I tried to see whether the monkeys' performance could be improved was to run the test in the mornings as the monkey was tested in the afternoons in normal situations. The number of the data acquisition of the ones being tested in the mornings in Monkey1 was 255 sessions in total. These sessions were tested with different versions of the touchscreen task such as 3 shapes or 2 shapes. However, these sessions tested at different times of the day did not make any difference with relation to the monkey's performance as the monkey was still performing at chance in most of the time. Additionally, the percent correct of breach trials is close to the one that when the monkey was tested in the afternoons. I have also tried to collect the data in Monkey 2 with testing her in the mornings, however, due to the individual differences between the monkeys, M2 did not get involved in the touchscreen task while being tested in the mornings. This might be explained by the fact that there would often be a home unit cleaning procedure in the colony in the early mornings on some days, which might have caused some effects to M2 being willing to perform on the touchscreen. Additionally, there were some issues with M2 in terms of the mental health as M2 was found a bit depressed when looking at the behaviour. Considering of the animal welfare, M2 was tested in the afternoons only. Moreover, as mentioned before, due to the individual differences, M2 was given a short delay between trials when performing the touchscreen task.

Figure 3.19. shows the percent correct for breach trials when M1 was tested in the mornings. These sessions were tested with 2 visual stimuli and the chance level is 50%. As can be seen from the figure, the performance stays stable across sessions without significant improvement (paired t-test, t(9) = 0.6806, *p* = 0.5133). In parallel, the percent correct does not show obvious decrease. This suggests that either testing in the mornings or in the afternoons does not make many differences in terms of improving the monkey's performance.

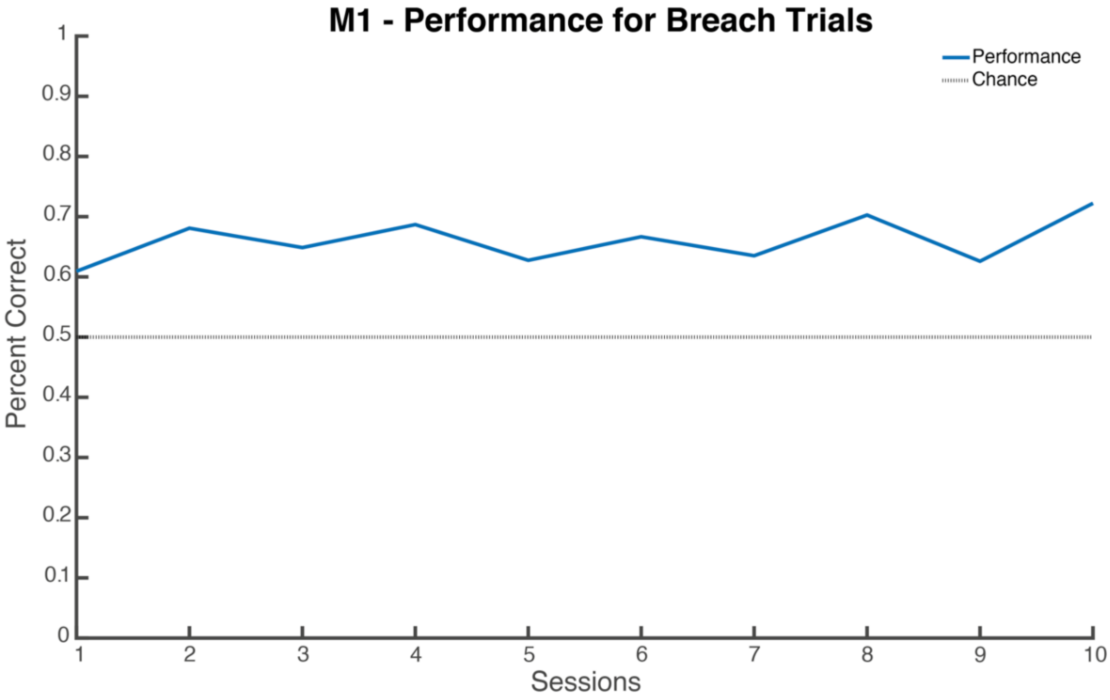


Figure 3.19. M1 – Performance for Breach Trials. Sessions tested with 2 visual stimuli, and the chance level here is 50%. The blue line shows the performance, and the dotted line indicates

the chance level based on the number of visual stimuli presented within the testing sessions (paired t-test, N = 20, t(9) = 0.6806, $p = 0.5133$).

Moreover, we also investigated the percent correct by stimulus type when tested the monkey in the mornings. The figure below (Figure 3.20.) shows the performance by shapes and colours for breach trials in M1. These sessions were the ones tested with 2 visual stimuli being presented on the screen. Additionally, the top left quadrant was blocked, meaning that the stimuli were not presented in that quadrant. As is shown in the figure, I tested whether the percent correct performance in shape and colour conditions differed. There was a significant difference in performance between the two in M1 (paired t-test, t(4) = 5.7305, $p = 0.0046$). This suggests that the monkey showed bias towards the colours when tested in the mornings with the top left quadrant blocked.

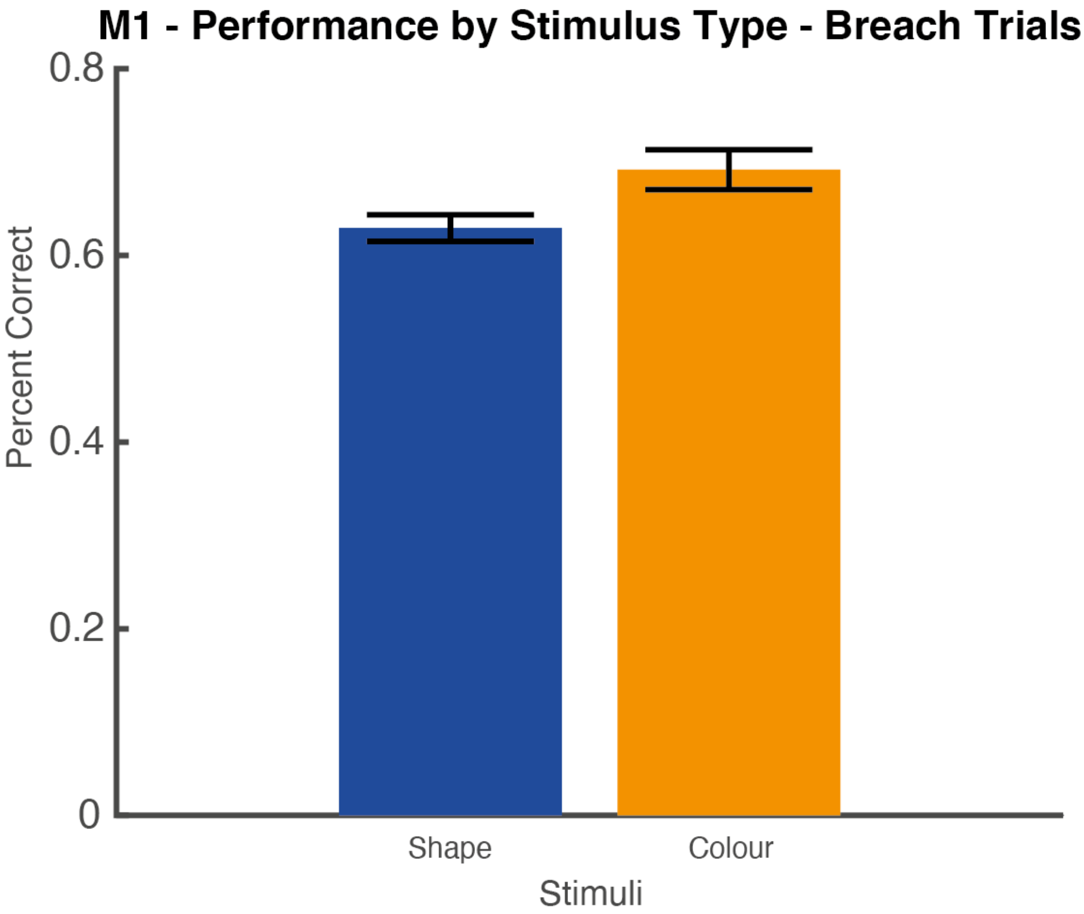


Figure 3.20. M1 – Performance by Stimulus Type - Breach Trials. Sessions tested with 2 visual stimuli, and the chance level is 50%. The percent correct for breach trials in shapes and colours

is presented in blue and orange, respectively (paired t-test, N = 6, t(4) = 5.7305, *p* = 0.0046, error bars are standard deviation).

3.8 Summary of Number of Sessions in the 3 Phases

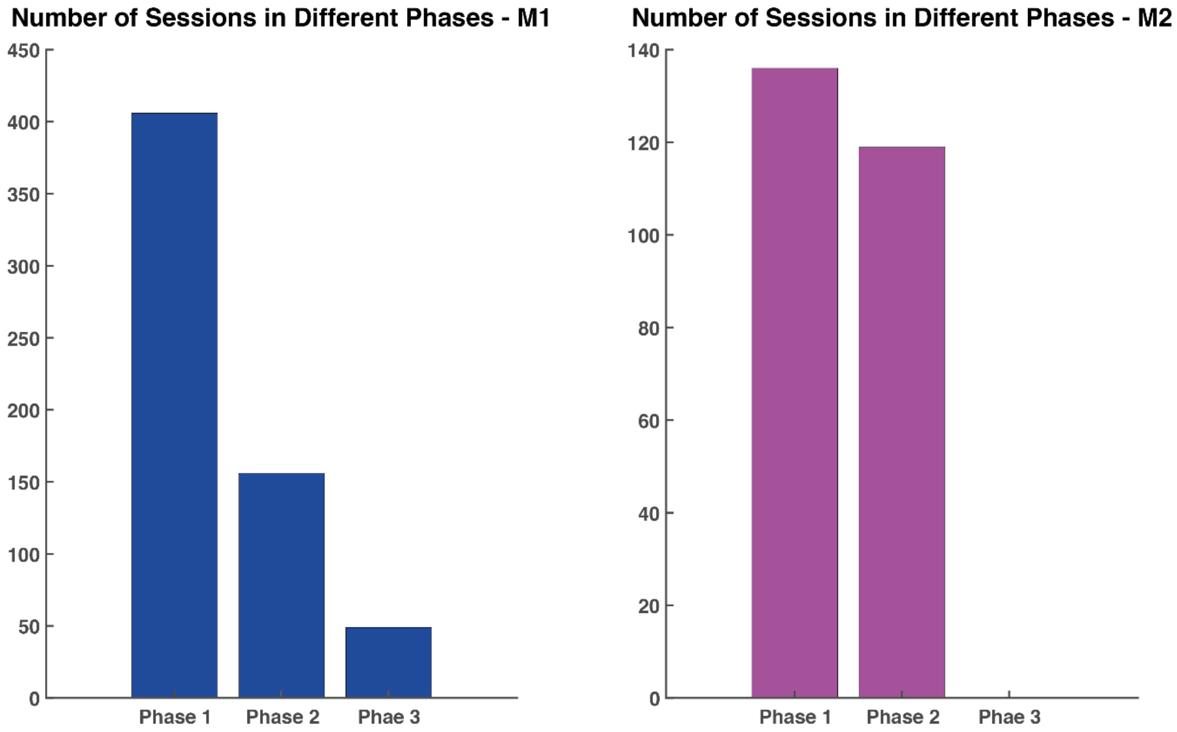


Figure 3.21. Summary of the number of sessions in different phases including Phase 1, 2 and 3. The panel on the left is the summary of sessions collected in M1 and the right panel is the summary of sessions collected in M2. No data of Phase 3 was applicable in M2.

As is shown in the figure above (Figure 3.21.), more sessions were collected in M1 compared to M2 among these 3 phases. Additionally, abundant sessions were needed for the monkeys to go through different phases. Moreover, some reasons for the large amount of data collection in Phase 1 and 2 are due to the COVID-19 pandemic necessitating me to re-train the monkeys as well as the necessary attempt to improve the touchscreen task and to increase the monkeys' performance.

The table below lists the data collected from different phases in monkey 1 and 2.

Table 3.1. Number of sessions collected in the two macaques

Number of Sessions	Monkey 1	Monkey 2
Phase 1	406	136
Phase 2	156	119
Phase 3	49	N/A

These sessions collected from different phases in both monkeys contain the ones tested daily and within the same day, meaning that there were some days that the monkeys were tested with different versions (shape/colour). On top of that, there were extensive data collected in both monkeys for Phase 1 and 2 due to the pandemic of COVID-19 as the monkeys needed to be trained to re-familiarize with the touchscreen task after a long period of the breakdown. Additionally, there were some sessions tested with minor changes of the scripts for improving the monkeys' performance, and details will be given in latter parts of the chapters. Monkey 1 was proceeded to the final phase which was the combinatorial learning while Monkey 2 was not proceeded to the final phase due to some unexpected issues. To be mentioned, M2 was tested almost one year behind M1, and M1 was started with the shape version while M2 was tested with the colour version first for counterbalance. In addition, only a few sessions were collected in M1 for the final phase due to the policy and regulated rules of the animal welfare.

The table below lists the detailed data collected from different phases in monkey 1 and 2.

Table 3.2. Number of sessions collected in the two macaques

Number of sessions	Monkey 1	Monkey 2
3 shapes/3sounds	62	N/A
6 shapes/3sounds	262	N/A
2 shapes/2 sounds	200	111
3 colours/3 sounds	68	62
2 colours/2 sounds	183	82

This table lists the data collected from both monkeys in different phases in more details. To be mentioned, these numbers indicate the testing sessions including different numbers of colours or shapes in Phase 1 or 2, therefore, the total number varies from the sum of all sessions as some of the sessions were overlapped. Rows 1-5 are the different initial testing phases. These are either Phase 1 (N = 406 for M1; 136 for M2) where either colour or shape was tested for at least a week, or Phase 2 (N = 156 for M1; 119 for M2) switching between Colour and Shape, sometimes in the same day.)

3.9 Discussion

I implemented numerous approaches to optimise the monkeys' performance including optimising reward and time outs, adjusting the monkeys' preferred touching positions and testing in different time periods such as mornings and afternoons. Additionally, I introduced the foil stimuli to the monkeys to assess whether the monkey could learn to ignore the foil stimuli. However, none of these helped with substantially increasing the monkeys' performance.

The number of trials tested in Phase 3 decreased when compared to Phase 1 and 2. One possibility could be that the complexity of the touchscreen task was increased as Phase 3 was tested with the combination of colours and shapes.

Apart from that, the performance of the monkeys could be easily affected by conditions in the colony. For example, people walking by, noise in the colony and other monkeys being taken out, therefore, not all trials would be meaningful or valuable of the testing sessions. Although the experimental conditions might not be perfect, I tried to pick the testing time of the day where the monkey performed as well as it could and there were fewer distractions.

Moreover, changes of parameters even if a small change may have effects on the task as well, suggesting that there might be a section among the parameters that would be the most suitable to the monkeys while performing the task. I therefore optimised the testing of the two animals to ensure they performed a relatively high number of trials for each session, and the computer also monitored performance online giving a time-out when performance fell during the testing session. In terms of the visual and auditory stimuli we generated, the monkey seemed to learn the associations fairly well and showed above-chance breaches throughout the testing trials. I also tested the monkeys on several additional foil stimuli incorporated on the touch screen. But the monkeys, interestingly, quickly learned to reject the visual foils. Therefore, I dispensed with using the visual foils and focused on training the monkeys on the sound associations with the colours or the shapes.

As mentioned before, one of the aspects having a detrimental impact on the touchscreen task or the monkeys' performance might be the experimental environment. Laboratory training may have been less distracting, and the monkeys may have performed better in the laboratory, but the numbers of trials that were obtained in the colony with the monkey pacing itself would have been very difficult to achieve in the laboratory. There has been abundant data collection from animal studies in relation to cognitive abilities conducted in zoos over the last few decades (Egelkamp & Ross, 2019), and many of these zoo-based studies were progressed even with the animals residing still among the monkeys' own social groups (Gazes et al., 2019; Hopper, 2017). Additionally, there were some studies conducted on marmosets with characteristically designed 'testing box' (Calapai et al., 2022) or similar home-cage apparatus (Nakamura et al., 2018), ensuring a quiet experimental environment during the testing. Some of these smaller testing apparatuses such as the ones applied to marmosets might be beneficial to the tasks that researchers conducted as the impact of any other noises was limited. Additionally, the animals could be more focused rather than distracted by anything going on in the colony or any other animals in the same unit.

I have tried to make the experimental environment as optimal as possible, opting to test the monkey in an individual testing home-cage in the corner of the colony away from much of the noise and distractions. By doing this, the monkey being tested could have a lower possibility to interact with other monkeys in the same colony. Moreover, the impact was minimised when people walked around as the cage that the monkey being tested was at the end of the corridor. Apart from that, the amplifier was connected to the laptop during the test for adjusting the volume to ensure the sound being played could be heard clearly by the monkeys. By doing all these, the impact of testing in the colony was minimised to ensure as optimal of an experimental environment during each testing session as possible given the nature of testing in the colony.

Another possibility that might have impact on the monkeys' performance was the straight bars of the home-cage as I mentioned formerly. Due to the limited access to the touchscreen device, the monkeys' behaviour might also be affected when reaching to the stimulus presented not close to where they were able to touch. In this case, they might give me lower performance than they could. Otherwise, it is worth thinking about the ways to improve the touchscreen device and training, meaning that it could be easier for the monkeys to touch the screen without any limitations of the straight bar of the home-cage, but this was not possible to further improve within the timeframe of this work. That said, I am confident having tried several ways to improve the monkeys' behaviour that optimising the paradigm individually for each monkey gave consistent behavioural performance. The exception was when I gave M2 too long of a time out and it would start to give up performing on the task, by giving too few trials and disengaging from the task. The other exception was when I blocked the quadrants that the monkeys preferred to touch, which gave chance performance for several days.

To conclude, there are many factors that could cause the monkeys not to perform optimally, but having tested different ways to improve performance I am confident that the task difficulty was the main reason the monkeys did not perform at high levels throughout the testing sessions. That said it is remarkable that they did perform well above chance, they just could not sustain it throughout the testing sessions.

Chapter 4: Switching Between Colour and Shape Versions

Abstract

Understanding the mechanism underlying combinatorial semantics is important. The touchscreen paradigm we established allows us to investigate whether the monkey can identify an object by hearing sounds that define its individual or integrated shape and colour properties. Prior to proceed the touchscreen testing to final combinatorial learning phase, it is important to understand whether the monkey can alternate between colour and shape learning, and whether they show a bias to either the shapes or colours to avoid any possibilities that the monkey would just rely on one of the properties to perform the combinatorial learning task. Therefore, we progressed the second phase with alternation between the shape and colour testing versions with several steps to minimise the possibility of any bias towards the shapes or colours. In specific, the monkey was firstly tested with the alternation between weeks, followed by the switching between days, and finally the colour and shape alternation within the same day after the monkey met the set criteria. Although the results from Phase 1 training led me to expect a bias in the colour or shape object performance, the results from Phase 2 indicate that the monkey did not show much bias to the shapes or colours ahead of being tested on the final combinatorial learning phase.

4.1 Introduction

Studies have been conducted in relation to whether monkeys are able to recognize visual objects by shapes or colours. A study conducted in early 1990s exploring how the monkeys recognize and discriminate the other monkeys. This study found that monkeys could discriminate one monkey from the others when presenting the stimuli of line drawings with merely restriction to other monkeys during the test (Dittrich, 1994). On top of that, Rinne et al. performed an audio-visual selective attention task in both humans and monkeys indicating that the monkeys are capable to conduct the audio-visual selective attention task, which also provides insights into comparative human and NHPs cognition (Rinne, Muers, Salo, Slater, & Petkov, 2017). However, Rinne et al. also noted that it took two years for the monkeys to perform at high levels (above chance) on the task, and they could not sustain it for every testing session, which is similar to my impression here. Moreover, Evans et al. conducted a auditory-visual recognition task showing cross-modal perception in NHPs (Evans, Howell, & Westergaard, 2005).

Apart from that, Vázquez et al. conducted a comparative study on humans and monkeys exploring whether the monkeys can discriminate the orientation of visual lines (Vázquez, Cano, & Acuña, 2000), the results indicated that the monkeys are capable to discriminate the line with different orientations, however, taking longer time compared to humans. Another study focusing on auditory discrimination found that monkeys and rats could discriminate the tunes with distinguished intensity and octave transformations in sounds (D'Amato & Salmon, 1982) .

Moreover, a comparative study exploring the dimension of visual objects in relation to working memory conducted by Fehring et al. found significant bias towards either shape or colour properties in both humans and monkeys (Fehring et al., 2022). More specifically, they conducted similar match-to-sample experiments in humans and monkeys with minor differences of the experimental design such as the timing, duration of the matching period. The monkeys were given longer matching period compared to humans, and then a fixation spot was presented on the screen before the stimulus was presented. Additionally, the delay before the decision was made was varied in different timings. According to the behavioural results they found, the monkeys seemed to have a short response time to shapes compared to colours, and the accuracy of their responses to shapes was higher than to colours. However, humans showed opposite results in terms of the response time as well as the accuracy of either shapes or colours. The results indicated that humans had a longer response time to shapes, and the accuracy of the shapes is significantly lower compared to the colours. In summary, the results from humans and the monkeys suggest that the monkeys seem to have a bias towards the shapes, and humans show the bias to the colours.

Apart from that, a study conducting the Wisconsin Card Sorting Test (WCST) indicated that the monkeys showed the shape bias opposite in effect to humans which showed a colour preference (Mansouri, Buckley, Fehring, & Tanaka, 2020). Note that these effects are opposite to those reported by Fehring et al., 2022 (Fehring et al., 2022), indicating that there is not a specific colour/shape bias by species, that these biases can differ potentially depending on the task or stimulus properties. Additionally, there was a Stop Signal Reaction Time (SSRT) task conducted in both monkeys and humans demonstrated that the shape bias occurred in monkeys while humans showed the colour bias (Ghasemian et al., 2021b). These are also consistent with the studies in the manual selection task (Kerzel & Schönhammer, 2013) as well as the colour and orientation task (Found & Müller, 1996).

With regards to data in apes and children, Vonk and Rastogi conducted a comparative touchscreen task between children and some other apes, finding that children have a bias towards the shape compared to the colour, however, the other apes involved in their study show a bias to the colour instead (Vonk & Rastogi, 2019). Categorising objects for information organization is human nature as humans are said to have naturally, and children prefer shapes rather than other features (Vonk & Rastogi, 2019). More specifically, when a novel object is presented to the children with a label, it is highly possible that they extend this to other objects with similar or same shapes rather than objects without sharing the same shape feature.

These findings across various cognitive studies raise some interesting questions to the study I conducted whether monkeys have a discrimination bias to shapes or colours, and whether distinct visual stimuli could have any impact on their behavioural performance. However, a major difference between these studies and my work is that in my task, I used the auditory stimuli and the number of the auditory stimuli in specific associations with the target visual stimuli which are used in the touchscreen task.

4.2 M1 – Switching Week by Week on 3 Colours/Shapes

As is shown in the figure (Figure 4.1.), M1 was tested with the alternation between 3 shapes and colours of the touchscreen task week by week, meaning that only shape version or colour version was tested within the week. The panel on the top indicates the number of breach differences in each session relative to the CI. Specifically, a permutation test shuffling the monkey's own response for 10 thousand times was conducted to calculate the number of breaches above 95% CI and the ones below the 5% CI, meaning that the breaches were calculated when the monkey breached above the CI or dropped below the CI while performing the task. Additionally, the number of breach difference was calculated with the number of breaches above 95% CI minus the number of breaches below 5% CI. The number of breach difference in negative values indicates that the monkey had many times that dropped below the 5% CI, meaning that the performance was even much lower than the chance. As mentioned before, to determine whether the monkey was able to further proceed with the touchscreen task, a threshold was set to 10 for the number of breach differences. If the monkey could make the majority of consistent significant sessions within two weeks period, it would suggest that the monkey could be able to proceed further on the task. Practically however, even if the monkey showed it was ready to move to the next phase, often to maintain a stable performance, the monkey was tested with more sessions than were actually needed.

Sessions that reach the threshold are marked in orange stars in the below figures and it can be seen from that figure that the monkey was able to reach the threshold within a testing week. In parallel, the percent correct of breach trials is presented in the panel below. These breach trials are the ones when performance exceeds the 95% CI, and the percent correct of these trials are calculated. It can be found that the monkey could perform well in some of the sessions.

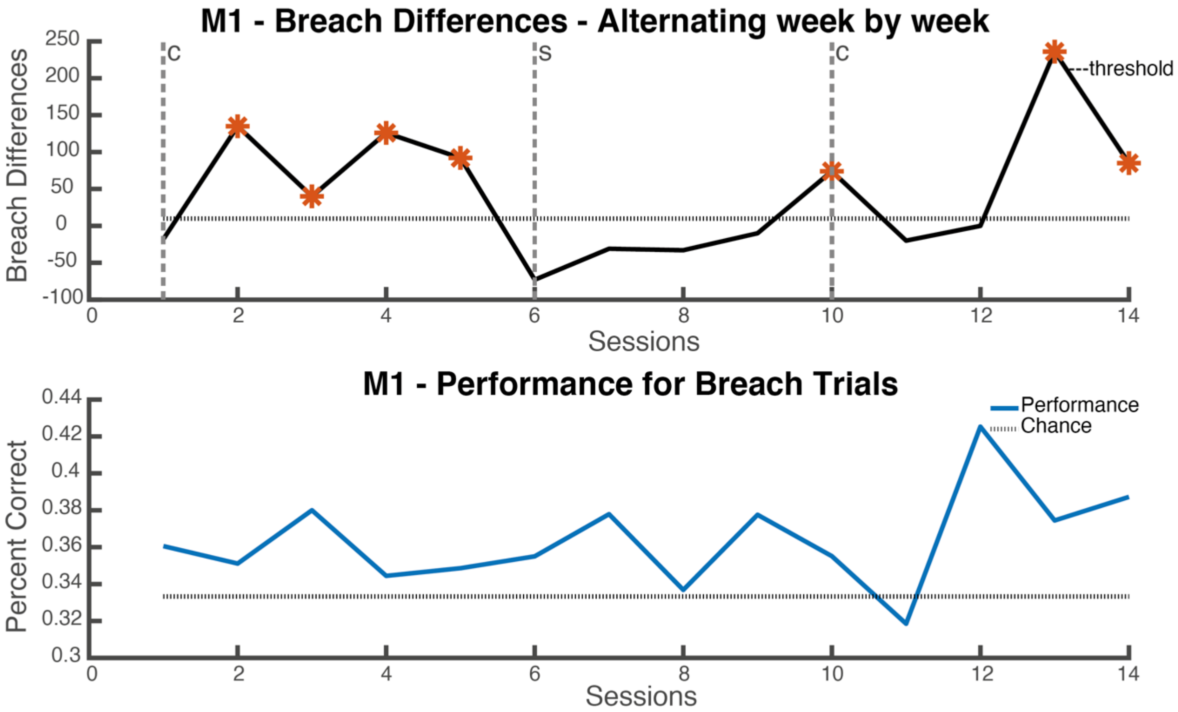


Figure 4.1. M1 – Breach Difference – Alternating week by week. The solid line in black in the top panel indicates the breach differences relative to the confidence intervals by calculating the number of breaches above 95% CI minus the number of breaches below 5% CI. These are the breaches that the monkey breaches above the distributed areas within the confidence intervals. The orange stars represent the significant sessions that the monkey reaches the threshold, and the threshold is 10 shown in the horizontal dotted line. C and S are colour and shape versions, meaning the start of the testing sessions with colours or shapes. The bottom panel indicates the percent correct for breach trials within each session. The blue line shows the performance, and the dotted line indicates the chance level based on the stimuli presented during the testing session.

After the monkey reached the threshold we set and necessary sessions for maintaining the stable performance, the monkey could be tested further with alternating the colour and shape versions day by day.

4.3 M1 – Switching Day by Day on 3 Colours/Shapes

The top panel in Figure 4.2. shows the number of breach difference when alternating the shape and colour version day by day, meaning that one day for shape only testing, and the following day for colour only testing. Theses contiguous sessions were started with the shape only testing, as is shown in the figure, there are many sessions that M1 reached the threshold consistently. The monkey was still capable to perform the touchscreen task and meet the criterion with a higher frequency of alternation.

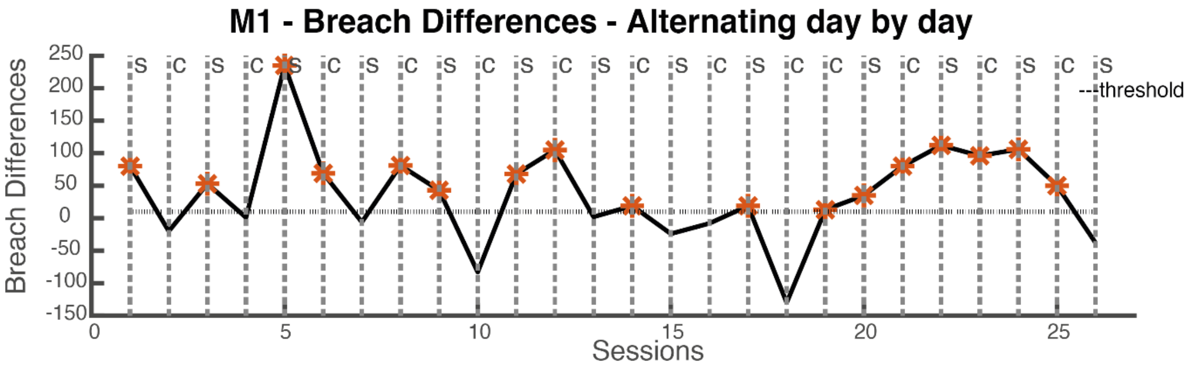


Figure 4.2. M1 – Breach Difference – Alternating day by day. The solid line in black in the top panel indicates the breach differences relative to the confidence intervals by calculating the number of breaches above 95% CI minus the number of breaches below 5% CI. These are the breaches that the monkey breaches above the distributed areas within the confidence intervals. The orange stars represent the significant sessions that the monkey reaches the threshold, and the threshold is 10 shown in the horizontal dotted line. C and S are colour and shape versions, meaning the start of the testing sessions with colours or shapes.

There is slight difference among the colours and shapes while switching the different testing versions day by day as is shown in the figure (Figure 4.3.). However, the percent correct for breach trials in shape only and colour only versions is slightly higher compared to switching week by week. Apart from that, the monkey seems still not to select the colour purple among the different colours. In comparison with the percent correct of blue or orange, the percent correct of the colour purple is lower (ANOVA test, F(2,36) = 0.32, *p* = 0.7298). The result in Figure 4.3. does not show many differences among the 3 shapes (ANOVA test, F(2,36) = 1.95, *p* = 0.1570). In other words, there seems to be no bias towards a particular visual stimulus.

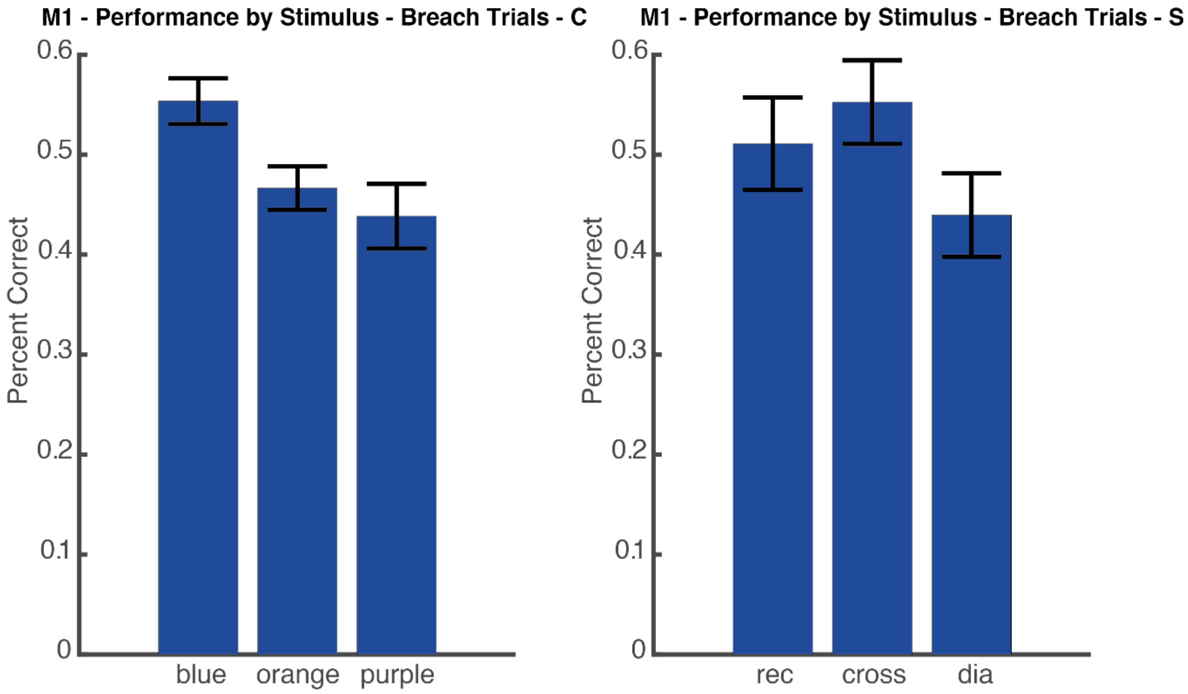


Figure 4.3. M1 – Performance by Stimulus – Breach Trials. The left panel shows the performance by stimulus on colours and the right panel presents the performance by stimulus on shapes, and the percent correct is only showing the breach trials (rec: rectangle, dia: diamond. Left panel: N = 13, $p = 0.7298$, Total degrees of freedom: 38, Error Degrees of Freedom: 36. Right panel: N = 13, $p = 0.1570$, Total degrees of freedom: 38, Error Degrees of Freedom: 36, error bars are standard deviation).

Additionally, Figure 4.4. shows the performance by colour only and shape only when switching the testing versions day by day. As can be seen from the figure, the mean performance of shapes for breach trials is slightly lower than colours, however, the performance could reach to approximately 70% correct in some sessions. Apart from that, the result indicates that M1 does not show a significant bias (paired t-test, t(12) = 0.8375, $p = 0.4187$) towards shape or colour when switching the versions between colours and shapes day by day.

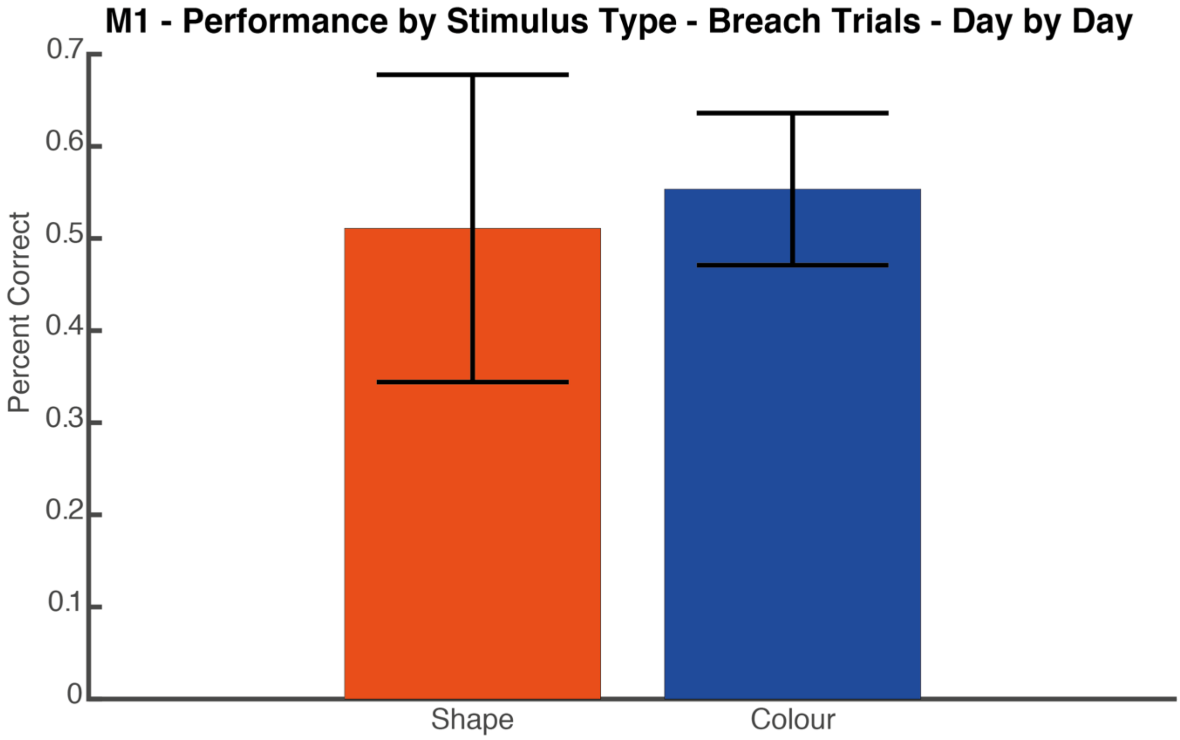


Figure 4.4. M1 – Performance by Stimulus Type – Breach Trials – Day by Day. These are sessions switching between the shape and colour versions day by day with 3 stimuli, and the performance by shape and colour for breach trials is presented (paired t-test, N = 26, t(12) = 0.8375, *p* = 0.4187, error bars are standard deviation).

As far as can be seen from the results above, there seems to be no significant difference between the stimulus being selected among the testing sessions, thus, the monkey was tested with switching the colour and shape within the same day after reached the threshold.

4.4 M1 – Switching within the Same Day on 3 Colours/Shapes

When switching between the colour and shape versions within the same day, the monkey was tested with either colour or shape first for an hour, then another hour was tested with the other version, meaning that the monkey was tested with two sessions each day. The top panel in Figure 4.5. shows the number of breach differences with alternation between the colours and shapes. These contiguous sessions started with the colour testing first, and we can see that there is a drop after switching to another version at the beginning. However, after some testing runs, the monkey was able to maintain the performance and reach the threshold. The panel at the bottom show that performance for breach trials, and the monkey could perform approximately 60% correct across the sessions. This might be explained by the reason that the monkey was tested on the shape and colour versions for a period which might have strengthen the impression and memory. Thus, it is possible that the monkey was able to perform the task with a higher performance.

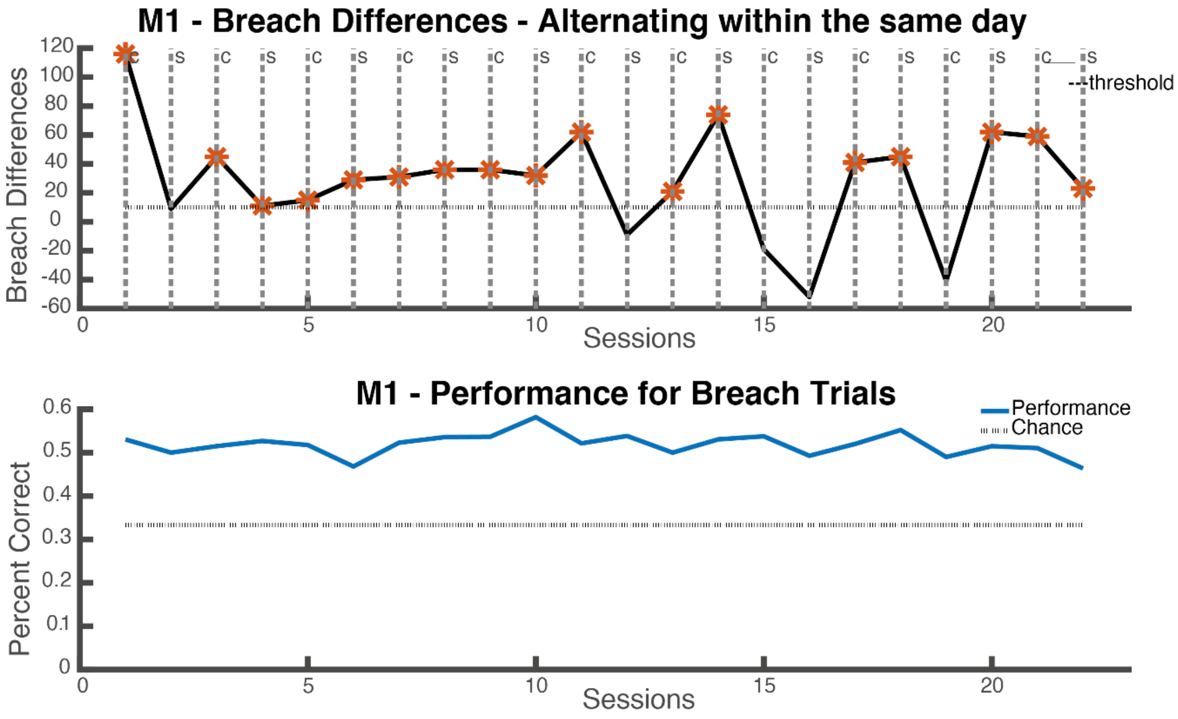


Figure 4.5. M1 – Breach Difference – Alternating within the same day. The solid line in black in the top panel indicates the breach differences relative to the confidence intervals by calculating the number of breaches above 95% CI minus the number of breaches below 5% CI. These are the breaches that the monkey breaches above the distributed areas within the confidence intervals. The orange stars represent the significant sessions that the monkey reaches the threshold, and the threshold is 10 shown in the horizontal dotted line. C and S are colour and shape versions, meaning the start of the testing sessions with colours or shapes. The bottom panel indicates the percent correct for breach trials within each session. The blue line shows the performance, and the dotted line indicates the chance level based on the stimuli presented during the testing session.

Figure 4.6. shows the result of the performance for each individual stimulus to see whether there is any bias. Similarly, M1 did not show any bias towards any of the shapes or colours (ANOVA test, left: F(2,30) = 2.09, *p* = 0.1411; right: F(2,30) = 0.38, *p* = 0.6894), and the percent correct by stimulus for breach trials does not vary much. However, it is still consistent with the results shown formerly when switching the colour and shape testing week by week or day by day that the colour purple was still not selected by M1 as many times as the other two.

In other words, the percent correct of the colour purple is lower than the other two colours as is shown in the figure.

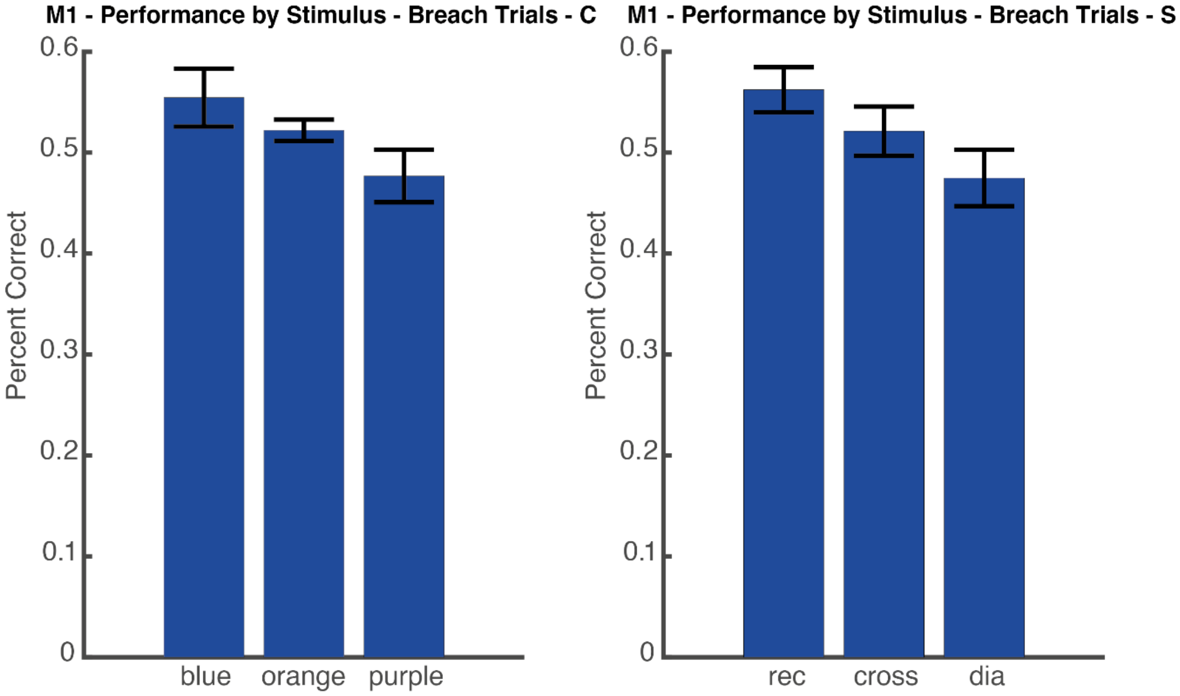


Figure 4.6. M1 – Performance by Stimulus – Breach Trials. The left panel shows the performance by stimulus on colours and the right panel presents the performance by stimulus on shapes, and the percent correct is only showing the breach trials (rec: rectangle, dia: diamond. Left panel: N = 11, *p* = 0.1411, Total degrees of freedom: 32, Error Degrees of Freedom: 30. Right panel: N = 11, *p* = 0.6894, Total degrees of freedom: 32, Error Degrees of Freedom: 30, error bars are standard deviation).

When alternating between the shape only and colour only versions within the same day, there is slightly increase in the percent correct of shapes and colours for breach trials as is shown in the figure (Figure 4.7.). Additionally, there seems to be no bias towards the shape or colour when switching the versions within the same day (paired t-test, t(10) = 0.1824, *p* = 0.8589). Apart from that, the percent correct for breach trials of shapes and colours slightly increased in comparison to switching the different versions week by week or day by day. This also suggests that the monkey could be tested further into the next phase.

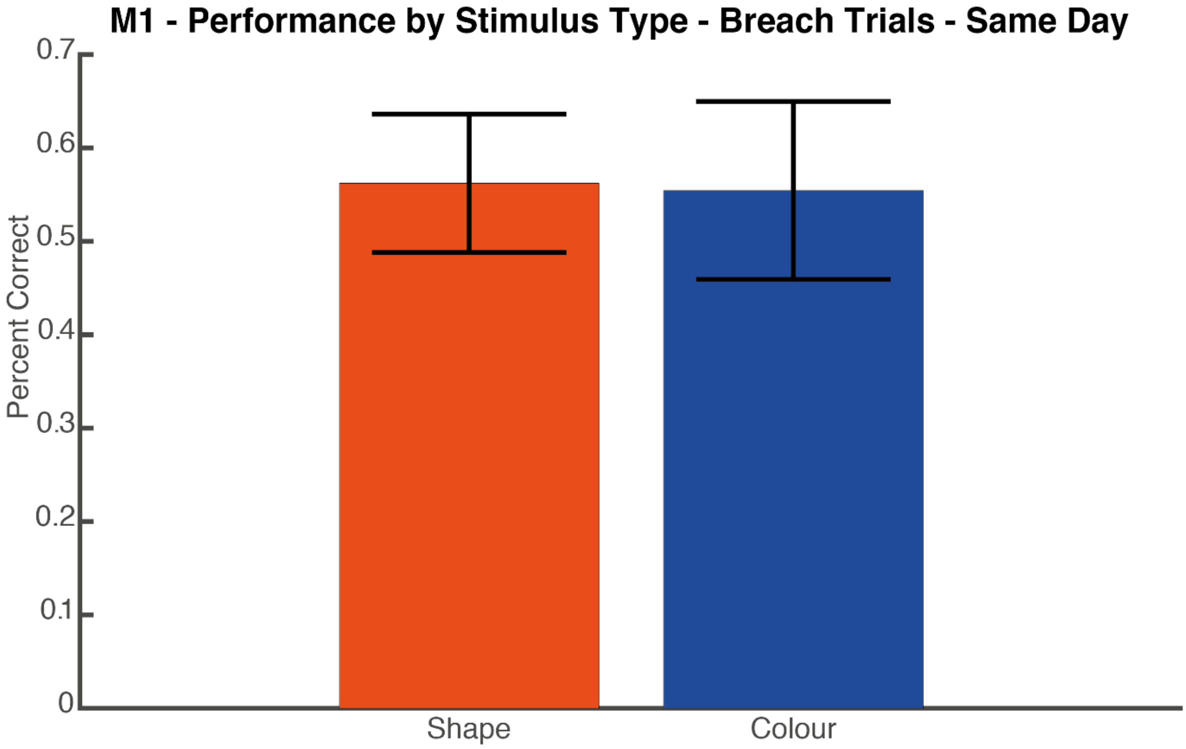


Figure 4.7. M1 – Performance by Stimulus Type – Breach Trials – Same Day. These are sessions switching between the shape and colour versions day by day with 3 stimuli, and the performance by shape and colour for breach trials is presented (paired t-test, N = 11, t(10) = 0.1824, $p = 0.8589$, error bars are standard deviation).

Furthermore, to examine whether the monkey learned the associations between the auditory and visual stimuli over time, I applied a linear regression model to estimate the how the number of breach differences changes over time. By doing this, it might be possible to see whether the monkey learned the associations and whether the performance was improved over time. More specifically, the linear regression model was applied into all the trials including shapes and colours, the shape only trials and the colour only trials when switching within the same day, and the sessions of the colour and shape only were the subset data of the ones in the left panel in the figure (Figure 4.8.). As is shown in the figure, the scatter plot in blue stars presents the distribution of the number of breach differences in each of the panel, there are only few ones lying on the straight line marked in red which indicates that the linear relationship between the number of breach differences and the time (sessions) is not closely related as can be found in these 3 panels. There are 22 sessions in total in the left panel, and the subset of 11 sessions of shape and colour only are plotted in the middle and the right panels. Overall, there is a decrease of the number of breach differences over time, and it can also be found in the colour only panel on the right side, but the number of breach differences of the shape stimulus slightly increased over time. Apart from that, none of them shows the significant increase of performance over time as the *p* values are not significant, indicating that performance was stable. This might be explained by the fact that the monkey might not be able to maintain the high performance all the time within one testing session, and that performance was stable for the monkey over time at this phase of training or that the performance could vary a lot in each of the individual sessions not showing a steady increase or decrease.

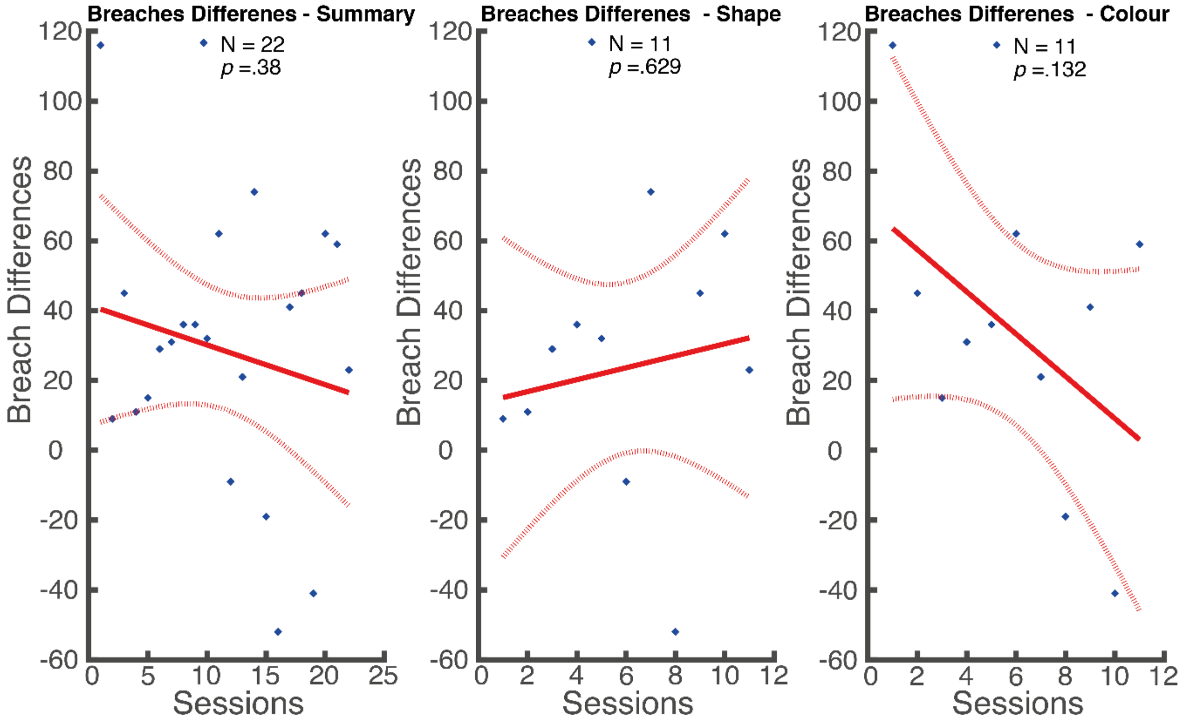


Figure 4.8. M1 – Linear Regression - Performance by Stimulus Type – Breach Difference. Blue stars are the distribution of the number of breach differences, and the red straight line indicates the liner relationship of the number of breach differences over time. N means the number of sessions, and p values are presented on the top of each panel indicating whether significant or not (left panel: error degrees of freedom: 20; root mean squared error: 37.8; middle panel: error degrees of freedom: 9; root mean squared error: 35.8; right panel: error degrees of freedom: 9; root mean squared error: 38.4).

As mentioned in previous chapters, the number of the stimuli was reduced from three to two in the end, and the results with 2 stimuli are presented as the followings.

4.5 M1 – Switching within the Same Day on 2 Colours/Shapes

The upper panel in Figure 4.9. shows the number of breach differences when alternating the colour and shape versions within the same day with two stimuli, meaning 2 shapes or 2 colours only in each testing session, and these sessions were the ones started with the shape testing first. As can be seen from the figure, the monkey did not perform well at the beginning, however, the performance was increased over more testing sessions. There are some sharp drops when switching from either colour or shape to the other, and the monkey could reach the threshold in consistent testing sessions. Additionally, the behavioural strategy of the monkey might have changed when performing the 2 stimuli touchscreen task. This is because the number of the stimuli was reduced to 2, making it easier for the monkey to select from the options. Moreover, the possibility of guessing when the monkey performing the task might have been increased as he just needed to pick one of the stimuli among 2. The chance of getting the juice reward has also been increased. Thus, the number of breach differences fluctuates a lot across the sessions at the very beginning of 2 stimuli testing. Additionally, the performance for breach trials increased compared to the 3 stimuli testing, which might suggest that 2 stimuli could be more sensible to be introduced in the touchscreen task as 3 could have been very difficult for the monkeys. Also, in some of the testing sessions, the performance could reach to approximately 80% correct. To be mentioned, there are some sessions that may not have enough breach trials which are presented in zeros. More specifically, the monkey did not do as many trials as the other sessions, which could be less than 50 in some cases.

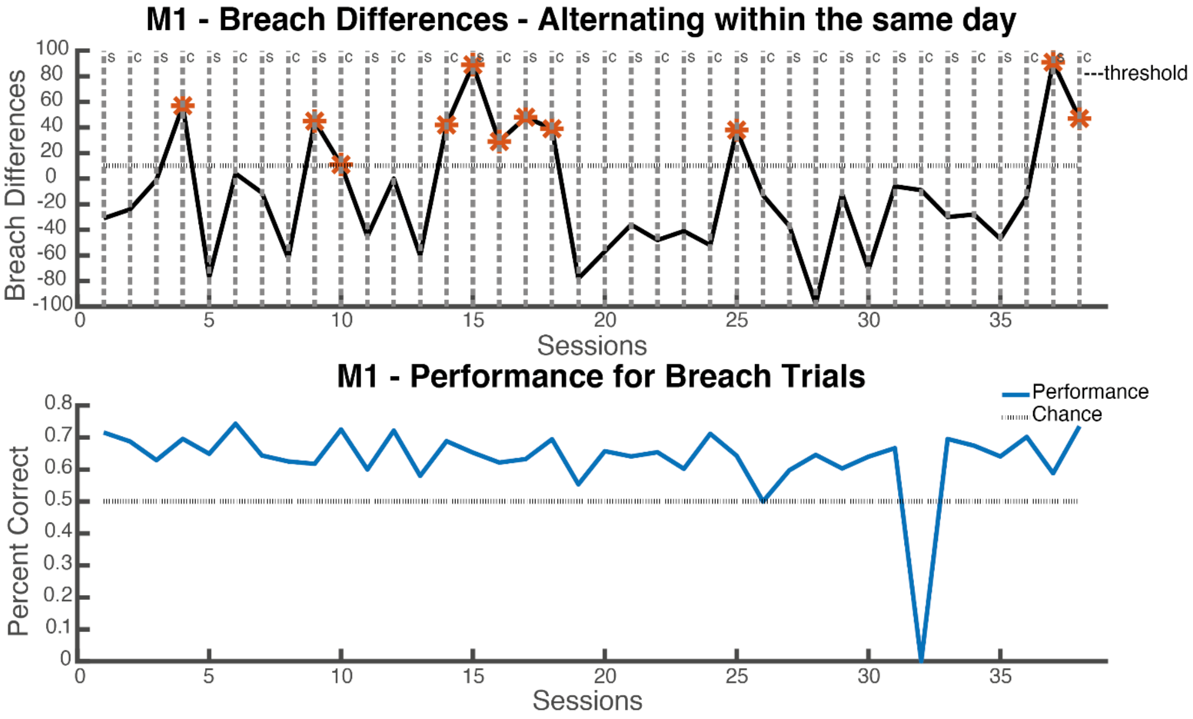


Figure 4.9. M1 – Breach Difference – Alternating within the same day. The solid line in black in the top panel indicates the breach differences relative to the confidence intervals by calculating the number of breaches above 95% CI minus the number of breaches below 5% CI. These are the breaches that the monkey breaches above the distributed areas within the confidence intervals. The orange stars represent the significant sessions that the monkey reaches the threshold, and the threshold is 10 shown in the horizontal dotted line. C and S are colour and shape versions, meaning the start of the testing sessions with colours or shapes. The bottom panel indicates the percent correct for breach trials within each session. The blue line shows the performance, and the dotted line indicates the chance level based on the stimuli presented during the testing session.

Similarly, it is worth trying to see whether the monkey shows a bias to a particular visual stimulus. And the colour purple as well as the shape rectangle were removed from the testing sessions. The reason for removing the purple rather than any of the other ones is that in the sessions tested early with 3 stimuli, the monkey did not select as many times as the other twos, and the percent correct of the colour purple was also lower compared to the other ones. Additionally, the rectangle was removed due to the reason that not only for counterbalance of the number with colour or shape property, but also that the other shapes are more unregular compared to rectangle. As can be seen from the figure (Figure 4.10.), the percent correct for breach trials in either shapes or colours increased compared to the testing sessions with 3 stimuli. There is not significant difference between the shapes of shape only sessions or the colours of the colour only sessions (paired t-test, left: t(18) = -1.3340, *p* = 0.1988; right: t(18) = 0.7378, *p* = 0.4702).

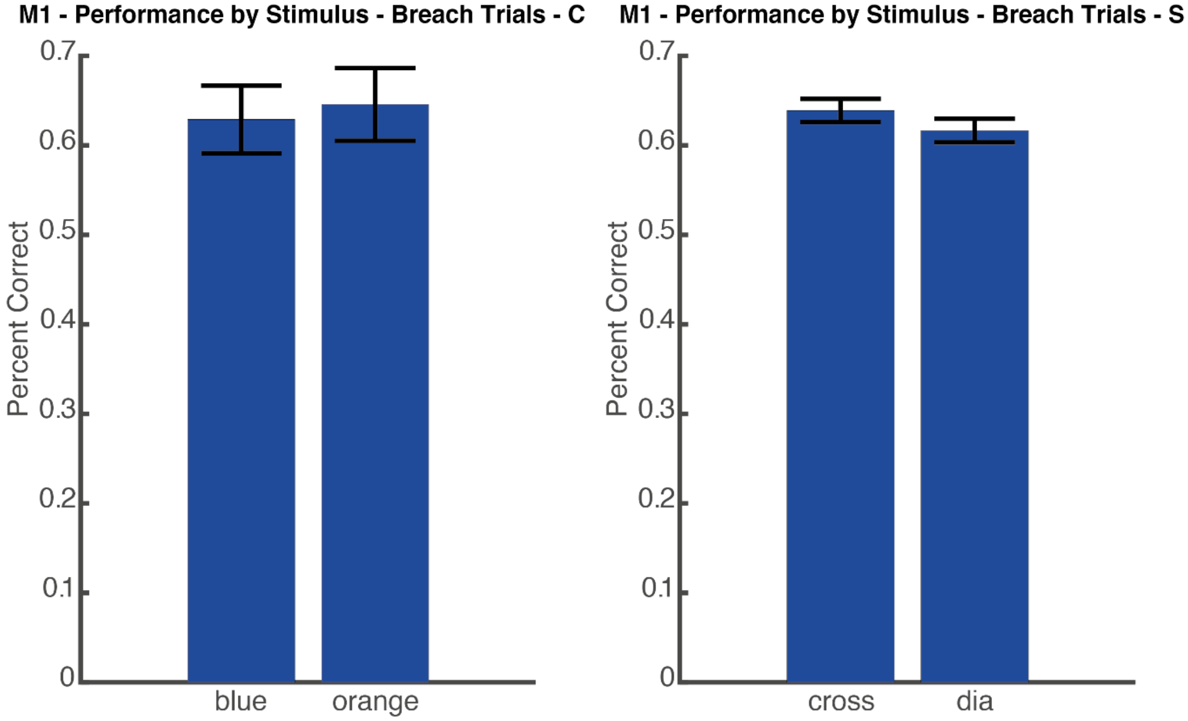


Figure 4.10. M1 – Performance by Stimulus – Breach Trials. The left panel shows the performance by stimulus on colours and the right panel presents the performance by stimulus on shapes, and the percent correct is only showing the breach trials (dia: diamond, paired t-test, N = 18, Left: t(18) = -1.3340, *p* = 0.1988; Right: t(18) = 0.7378, *p* = 0.4702, error bars are standard deviation).

To examine whether there is any bias towards the colour or shape property when tested with 2 stimuli in each of the sessions, a comparison between the shape and colour sessions was conducted as is shown in the figure below (Figure 4.11.). These sessions were contiguous sessions, and there is no significant difference between the shape or colour (paired t-test, t(18) = 0.2231, *p* = 0.8260), however, the percent correct for breach trials of both shapes and colours slightly increased compared to the testing sessions with 3 stimuli, reaching to 80% correct in some sessions.

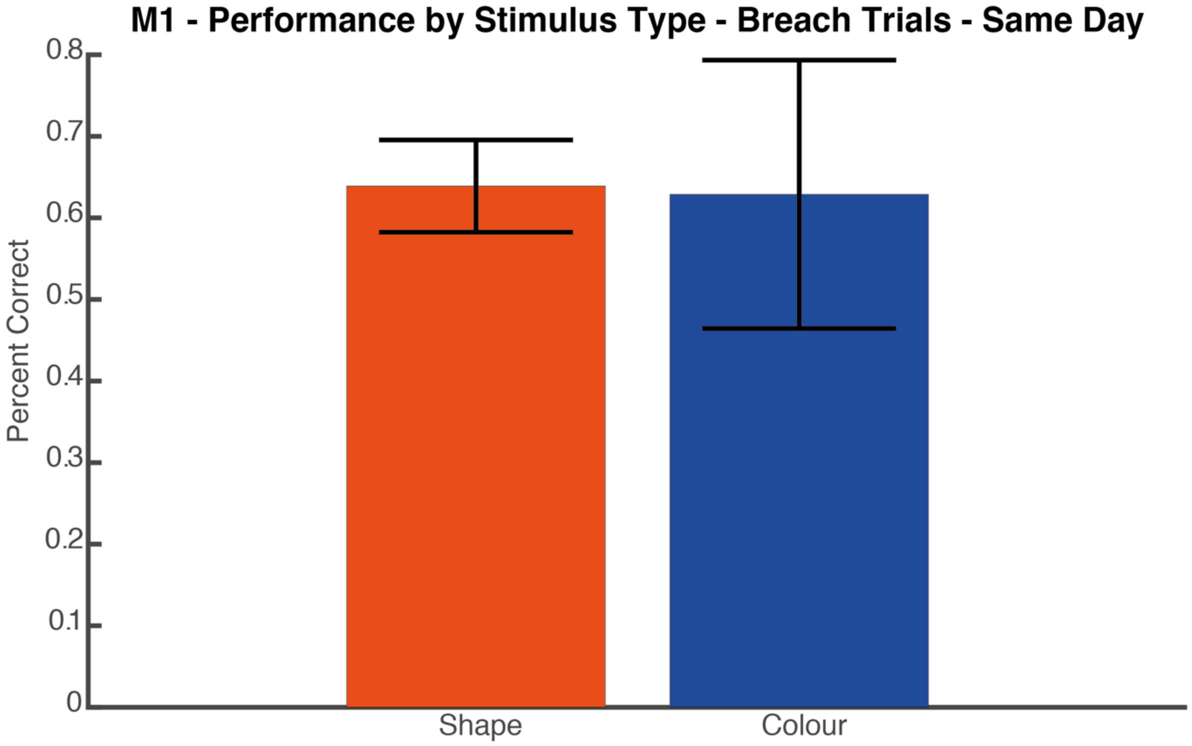


Figure 4.11. M1 – Performance by Stimulus Type – Breach Trials – Same Day. These are sessions switching between the shape and colour versions day by day with 2 stimuli, and the performance by shape and colour for breach trials did not significantly differ (paire t-test, N = 18, t(18) = 0.2231, *p* = 0.8260, error bars are standard deviation).

Moreover, to examine whether the monkey learned the associations between the auditory and visual stimuli over time when tested with 2 stimuli, a linear regression model was applied to estimate how the number of breach differences changes over time. As is shown in the figure (Figure 4.12.), the scatter plot in blue stars presents the distribution of the number of breach differences in each of the panel. It is similar to the result with 3 stimuli, only few ones of the number of breach differences in blue stars are lying on the straight line as marked in red, indicating that the linear relationship between the number of breach differences and the time (sessions) is not closely related as is shown in these 3 panels. There are 38 sessions in total in the left panel, and the subset of 19 sessions of shape and colour only are plotted in the middle and the right panels. Overall, there is a decrease of the number of breach differences over time, and it can also be found that the number of breach differences decreased in the colour only panel. However, the number of breach differences of the shape stimulus increased over time. Similarly, the *p* value in each of the panels is not significant, suggesting that there is not a strong relationship between the number of breach differences over time. In other words, the result does not show an increasing learning over time significantly. This might be due to the behavioural strategy change of the monkey as well as the difficulty of maintaining high performance all the time within a testing session, and the monkey might have plateaued in performance over time by this phase of testing or there might have been high variability across the testing sessions.

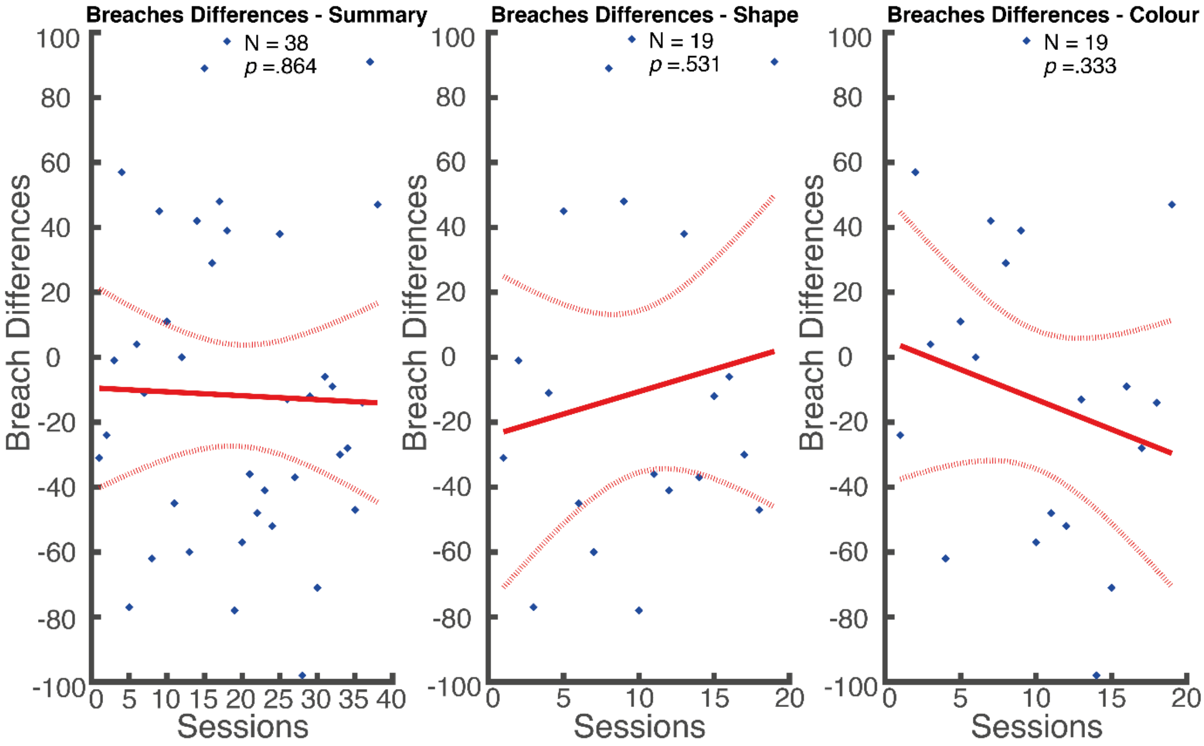


Figure 4.12. M1 – Linear Regression - Performance by Stimulus Type – Breach Difference. Blue stars are the distribution of the number of breach differences, and the red straight line indicates the liner relationship of the number of breach differences over time. N means the number of sessions, and p values are presented on the top of each panel indicating whether significant or not (left panel: error degrees of freedom: 36; root mean squared error: 47; middle panel: error degrees of freedom: 17; root mean squared error: 51; right panel: error degrees of freedom: 17; root mean squared error: 44.1).

4.6 M2 – Switching Day by Day on 2 Colours/Shapes

The second monkey was tested with the same stimuli used in M, and there were minor differences of the parameters in the testing scripts such as shorter timeout. This has been explained before as M2 would not be well involved in the touchscreen task if the timeout was set for a long time, not even to shut the screen down when the computation test was applied to improve the monkey's performance. The top panel in Figure 4.13. presents the number of breach differences under the condition of switching between colour and shape versions day by day. Compared to M1, there are fewer number of breach differences as M2 did not do as many trials as M1. Similarly, the monkey could not reach the threshold in all sessions, however, the monkey would be tested with the further step when met the criterion. The lower panel shows the performance for breach trials, and the percent correct is higher compared to M2 in average across the sessions. This might be because of that M1 could do many trials even more than one thousand, however, it did not mean that the more trials the monkey did, the higher percent correct it could be. This also indicates that the monkey can perform well across contiguous sessions.

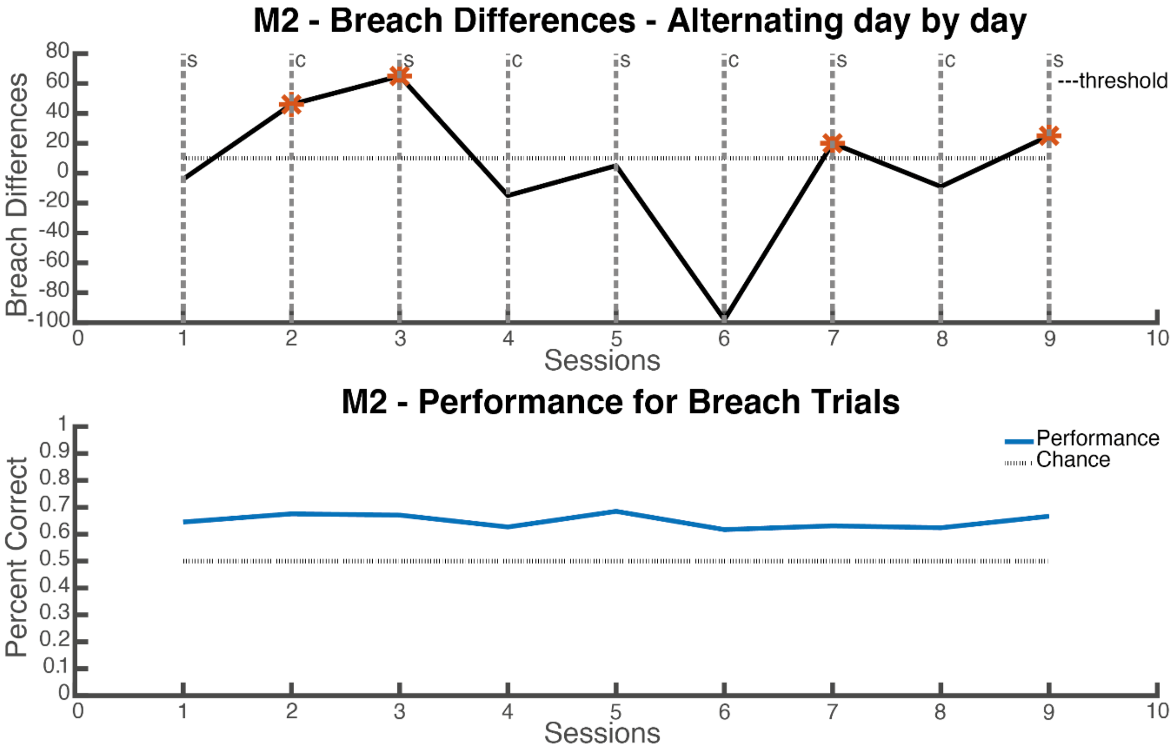


Figure 4.13. M2 – Breach Difference – Alternating day by day. The solid line in black in the top panel indicates the breach differences relative to the confidence intervals by calculating the number of breaches above 95% CI minus the number of breaches below 5% CI. These are the breaches that the monkey breaches above the distributed areas within the confidence intervals. The orange stars represent the significant sessions that the monkey reaches the threshold, and the threshold is 10 shown in the horizontal dotted line. C and S are colour and shape versions, meaning the start of the testing sessions with colours or shapes. The bottom panel indicates the percent correct for breach trials within each session. The blue line shows the performance, and the dotted line indicates the chance level based on the stimuli presented during the testing session.

Figure 4.14. shows the performance for breach trials by stimulus to see whether M2 shows the bias to any of the shapes or colours. The percent correct is approximately 60% in average among the two colours, and the percent correct is about 70% in two shapes. This is slightly higher when compared to M1. Additionally, there are not many differences between the two colours or the two shapes when the monkey needed to select one of them, indicating that there is not significant bias toward either the colour or the shapes (paired t-test, left: t(3) = 2.9298, *p* = 0.0610; right: t(4) = 0.1921, *p* = 0.8570).

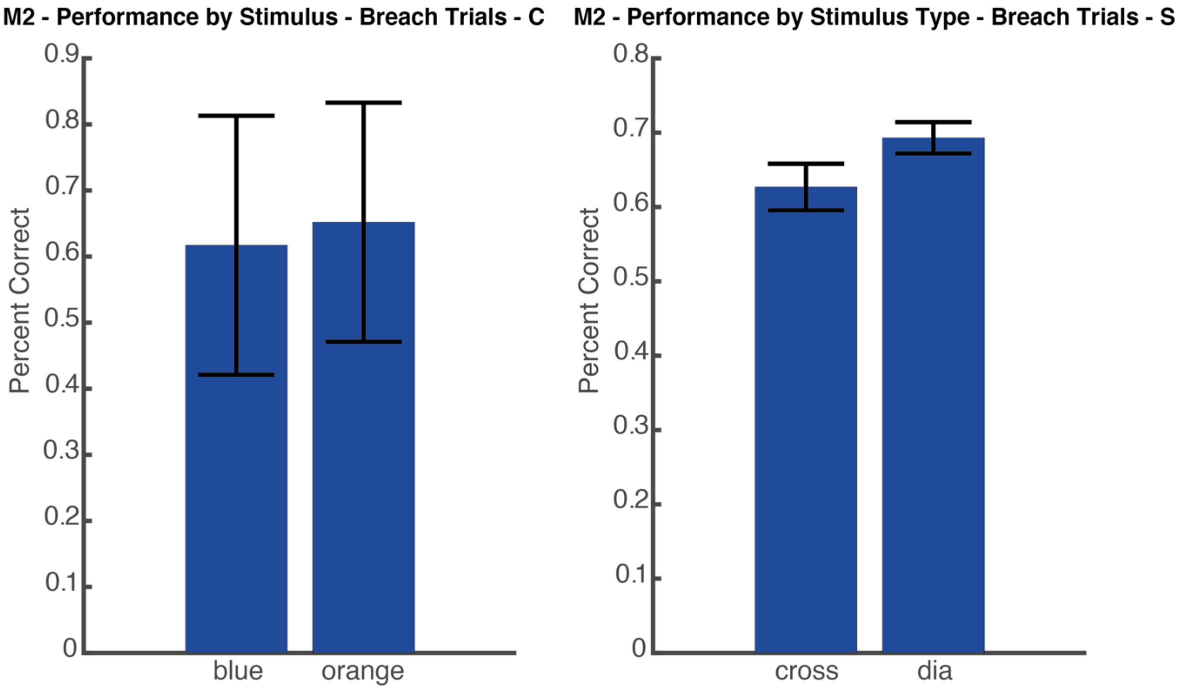


Figure 4.14. M2 – Performance by Stimulus – Breach Trials. The left panel shows the performance by stimulus on colours and the right panel presents the performance by stimulus on shapes, and the percent correct is only showing the breach trials. (dia: diamond, paired t-test, no significant differences in these results. Left: N = 4, t(3) = 2.9298, *p* = 0.0610; Right: N = 5, t(4) = 0.1921, *p* = 0.8570, error bars are standard deviation).

To examine whether M2 shows any bias towards the colour or shape property when tested with 2 stimuli, a comparison between the shape and colour sessions was conducted as is shown in the figure below (Figure 4.15.). The results show no significant difference (paired t-test, t(3) =

0.6561, *p* = 0.5586) between the shape or colour performance. Additionally, the percent correct for breach trials of both shapes and colours is approximately 60% which is similar to M1.

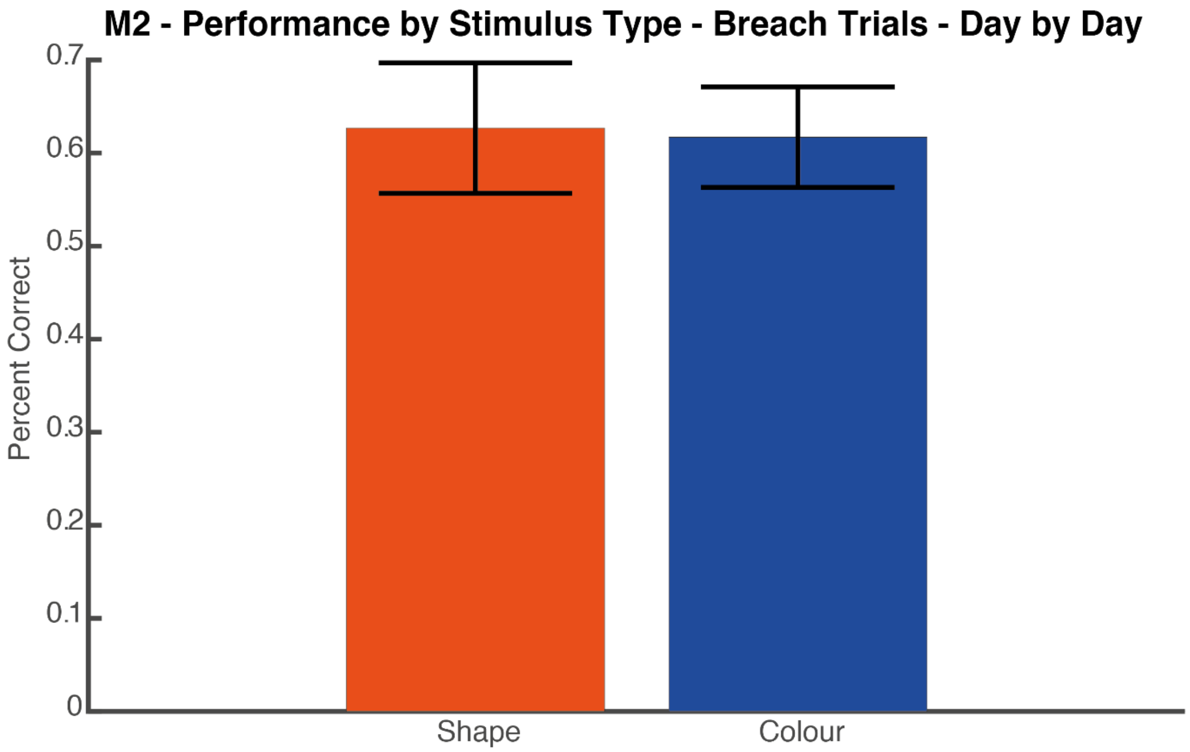


Figure 4.15. M2 – Performance by Stimulus Type – Breach Trials – Day by Day. These are sessions switching between the shape and colour versions day by day with 2 stimuli, and the performance by shape and colour for breach trials is presented (paired t-test, N = 9, t(3) = 0.6561, *p* = 0.5586, error bars are standard deviation).

4.7 M2 – Switching within the Same Day on 2 Colours/Shapes

The panel on the top in Figure 4.16. shows the number of breach differences when alternating the colour and shape versions within the same day with two stimuli, and these sessions were the ones started with the shape testing first. As is shown in the figure, the monkey did not perform well at the beginning, however, the number of breach differences increased across more testing sessions. Apart from that, the performance for breach trials could reach to approximately 85% correct. As mentioned in previous section, there are some sessions that may not have enough breach trials which are presented in zeros as the monkey might have performed fewer trials.

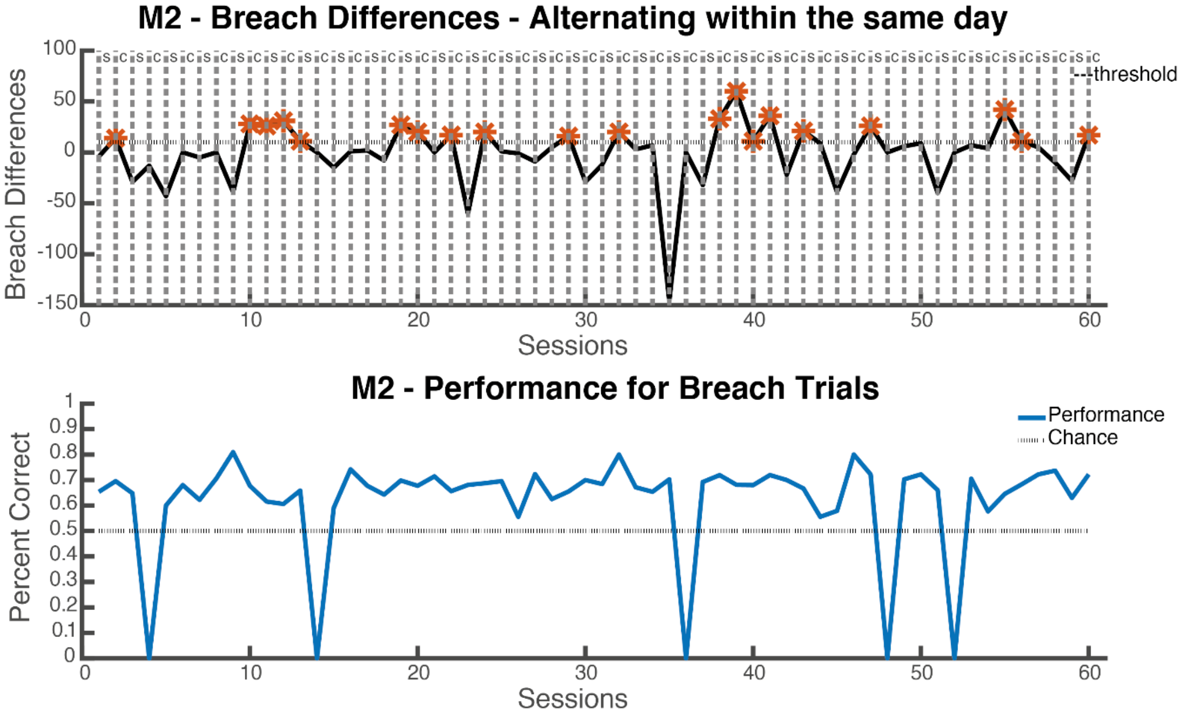


Figure 4.16. M2 – Breach Difference – Alternating within the same day. The solid line in black in the top panel indicates the breach differences relative to the confidence intervals by calculating the number of breaches above 95% CI minus the number of breaches below 5% CI. These are the breaches that the monkey breaches above the distributed areas within the confidence intervals. The orange stars represent the significant sessions that the monkey reaches the threshold, and the threshold is 10 shown in the horizontal dotted line. C and S are colour and shape versions, meaning the start of the testing sessions with colours or shapes. The bottom panel indicates the percent correct for breach trials within each session. The blue line shows the performance, and the dotted line indicates the chance level based on the stimuli presented during the testing session.

Figure 4.17. shows the performance for breach trials by stimulus to see whether M2 shows the bias to any of the shapes or colours when switching within the same day. As can be seen from the figure, there are not many differences between the either the two colours or the two shapes, indicating that there is no significant bias toward any of the visual stimuli. Figure 4.18. shows whether M2 has any bias towards the colour or shape property, and it can be found that the percent correct for breach trials of shapes is slightly higher than the colours.

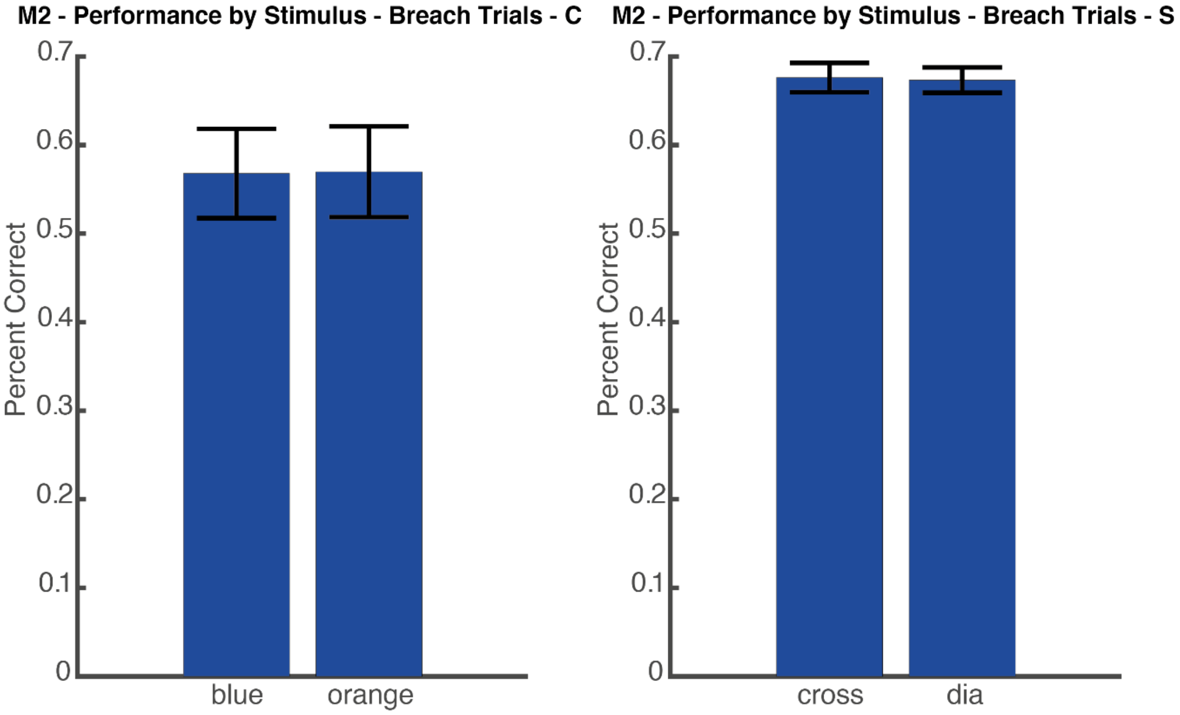


Figure 4.17. M2 – Performance by Stimulus – Breach Trials. The left panel shows the performance by stimulus on colours and the right panel presents the performance by stimulus on shapes, and the percent correct is only showing the breach trials (dia: diamond, paired t-test, N =30, Left: t (29) = 0.7874, *p* = 0.4375; Right: t(29) = -1.0542, *p* = 0.3005, error bars are standard deviation).

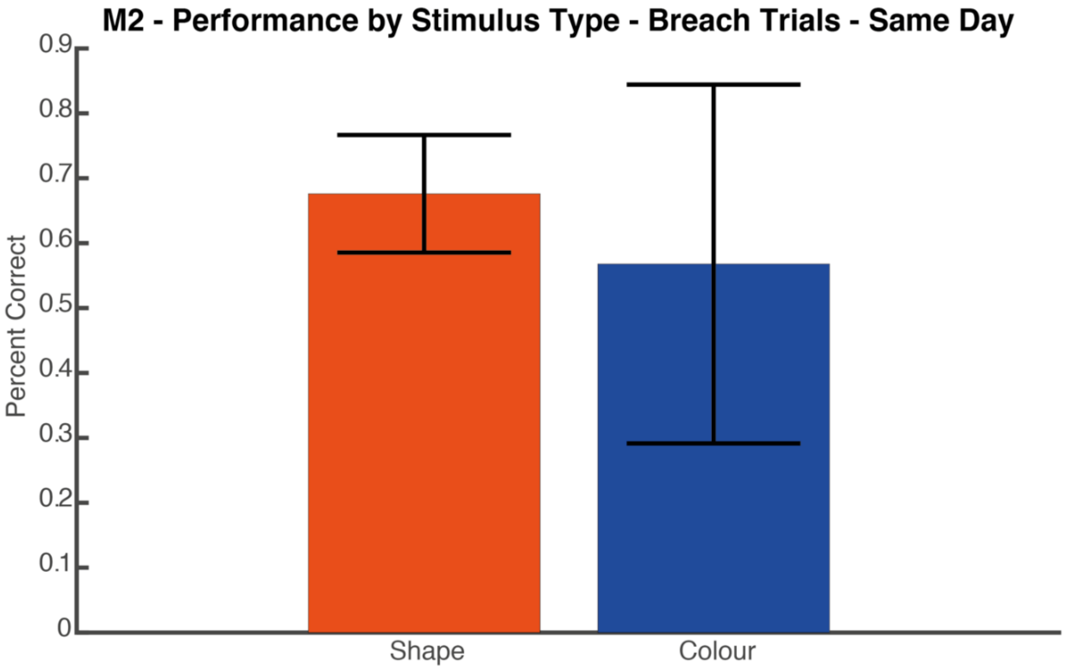


Figure 4.18. M2 – Performance by Stimulus Type – Breach Trials – Within the same day. These are sessions switching between the shape and colour versions day by day with 2 stimuli, and

the performance by shape and colour for breach trials is presented (paired t-test, N = 30, t(29) = 1.8271, *p* = 0.0780, error bars are standard deviation).

Furthermore, the linear regression model was utilised to examine whether M2 learned the associations between the audio-visual stimuli over time when tested with 2 objects. As can be seen from the figure (Figure 4.19.), the blue stars in each panel present the distribution of the number of breach difference, and few ones of the number of breach differences lye on the straight line as marked in red, suggesting that the linear relationship between the number of breach differences and the time (sessions) is not closely related. There are 60 sessions in total in the left panel, and the subset of 30 sessions of shape and colour only are plotted in the middle and the right panels. There is an increase of the number of breach differences over time as can be seen from the left and middle panels, and it can also be found that the number of breach differences in colour decreased as is shown in the right panel. Similarly, the *p* value in each of the panels is not significant, indicating that there is not an increasing learning over time.

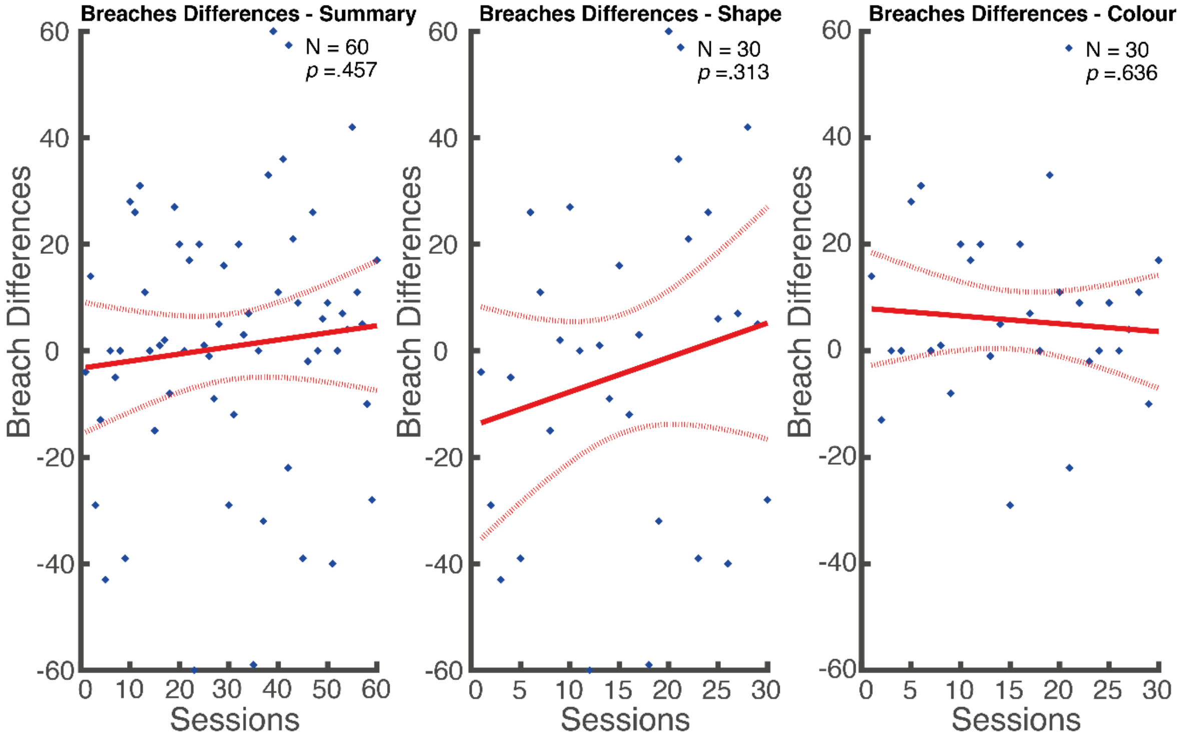


Figure 4.19. M2 – Linear Regression - Performance by Stimulus Type – Breach Difference. Blue stars are the distribution of the number of breach differences, and the red straight line indicates the liner relationship of the number of breach differences over time. N means the number of sessions, and p values are presented on the top of each panel indicating whether significant or not. (left panel: error degrees of freedom: 58; root mean squared error: 23.9; middle panel: error degrees of freedom: 28; root mean squared error: 29.8; right panel: error degrees of freedom: 28; root mean squared error: 14.6).

4.8 Discussion

Many studies conducted in relation to touchscreen learning used sounds to associate with visual objects (Calapai et al., 2022), or to use different labels referring to objects in black or white or in various colours (Medam & Fagot, 2016). Additionally, some of the studies have used distinguished shapes and colours in associations (Vonk & Rastogi, 2019), and many of the studies showed the shape bias in monkeys and the colour bias in humans (Fehring et al., 2022; Found & Müller, 1996; Ghasemian et al., 2021b; Kerzel & Schönhammer, 2013). There is also an interesting study conducting the comparative touchscreen task between children and some apes, they found that children have a bias towards the shape compared to the colour, however, the other apes involved in their study show a bias to the colour instead (Vonk & Rastogi, 2019). These studies indicate that NHPs may show bias to either the colours or the shapes.

There is no significant difference between the shapes and colours in the results that have been demonstrated. This is because when we found that when the monkey showed any bias to either the shape or the colour property, we implemented different ways for correction. More specifically, we have firstly tried to block the colour and shape trials within the same session, meaning that the number or percent of the trials in regards with the shape or colour property was blocked with a specific number or proportion. And the number of these two types of trials can be switched between each 20 or more trials according to the requirements might be needed. In another way, the proportion of shape and colour trials can be set to a particular number which can also be adjusted. Moreover, the number of the trial types can be adjusted, meaning that more colour or shape trials can be specified within a testing session. Similarly, the colour and shape trials can be randomised within a testing session as well to see whether any differences could occur compared to the trail types being blocked. By doing this, the bias towards the colours or shapes might be corrected before stepping into the final phase.

Apart from that, the reason for correcting any bias towards the shapes or colours is that to minimise the possibility of which that the monkey relies on one of the properties when tested with the combinatorial learning phase. Additionally, to make sure that the monkey learns the associations between the auditory stimuli and the visual stimuli, and the associations between the shape or colour properties of an object. And the results have indicated that the bias towards the colours or shapes was corrected as no significant difference is shown between the colours and shapes.

The alternation between the shape and colour versions of the touchscreen task is important as it could be difficult to know whether the monkey was selecting the integrated colour and shape objects relying on just one of the properties. If the monkey performed well on the shape only trials, then it might suggest that the monkey has just learned the associations between the sounds and the shapes. In contrast, the monkey might have just learned the sound and colour associations when the performance was significantly higher than the shapes. In this case, the data being collected in the final phase would not be as valuable as expected as this required the balanced learning of both the shapes and colours. Additionally, we switched the two different versions step by step with refinement. In specific, the monkey was firstly tested with the alternation for a period of at least one week on the shape version and switched to another version afterwards. Followed by the shorter time periods and higher frequency of switching to make sure that the monkey did not show any bias to the shapes or colours. Moreover, the methods and improvements in the experimental code make it possible to manage the number of trials in different properties which allows a good balance during the testing.

To summarise, the performance by stimulus shows slight differences among the shapes or colours when tested with 3-stimuli or 2-stimuli of the touchscreen task. Additionally, the results have also shown that the performance of the two types of trials in colour or shape do not have a significant difference, which is promising for the testing in the final phase that the monkey would not overly rely on either the shape or colour property.

Chapter 5: Combinatorial Learning Assessment in Macaque 1

Parts of this chapter have been presented as

Zhang. Z., Nacef. J., Poletiek. F., Wilson. B., Griffiths. T., Kikuchi. Y & Petkov. C. (2023) Evolutionary Prototype for Auditory Combinatorial Semantics in a Macaque. APAN2023: Advances and Perspectives in Auditory Neuroscience. Washington, DC. A72.

Zhang. Z., Nacef. J., Poletiek. F., Wilson. B., Griffiths. T., Kikuchi. Y & Petkov. C. (2023) Evolutionary Prototype for Auditory Combinatorial Semantics in a Macaque. *Society for Neuroscience* 2023. Washington, DC. PSTR412.01 / AA26.

Abstract

How human speech and language evolved from an auditory system shared with ancestors to living nonhuman animals remains an important open question, one with implications for the extent to which aspects of the human language system can be modelled in nonhuman animals. A key property of human language is combinatorial semantics, where information from a sequence of words is integrated to identify meaningful content. We designed a novel behavioural touchscreen task implemented with two rhesus monkeys (*Macaca mulatta*) in their home units. The task allowed us to study whether the monkeys could associate nonsense speech sounds with visual colours or shapes, prior to integrating the information contained in a sequence of two sounds identifying a specific object by its joint colour and shape properties. The paradigm was implemented in three key phases. In the first phase, the animals started by learning to associate the nonsense words with either specific colours or shapes. Followed by alternating the shape and colour versions in different frequencies. In the last phase, the monkeys need to learn to associate the sounds to an object by its colour and shape features. One of the monkeys progressed to the final phase of the experiment, where sequences of two sounds identified objects by both colour and shape properties. The macaque consistently chose the correct combinations whose joint colour and shape properties matched the informative content in the two sounds, over foil objects that contained one or none of the features referred to by the sounds (N = 49, $p < 0.005$). However, the first monkey was not able to generalise learning to probe trials of novel combinations. In summary, the results provide tentative support for a primate prototype of auditory combinatorial semantics in nonhuman primates, and the results also demonstrate that such abstract auditory learning is possible in at least one monkey but difficult for monkeys in general and potentially difficult to generalise to novel combinations, altogether providing novel insights into language evolution.

5.1 Introduction

One feature that is important to human language is combinatorial semantics. Language production can carry novel messages and humans can produce and understand the novel sentences which distinguishes human beings to other species, and it is considered as a unique feature in humans (Noam Chomsky & Skinner, 1959) . For example, these two words ‘red’ and ‘umbrella’, they have individual meanings when they are not integrated into a sequence which are the colour red and the object umbrella, respectively. However, when they are integrated into ‘red umbrella’, this refers to an umbrella that is red, meaning that an umbrella with its own colour, red. Otherwise, it could be an umbrella without knowing the colour which can be black or any other ones. Additionally, these two words are structured into a sequence by rule-based order, by which can be conveyed into meaningful sequences. Language can be systematically structured into a meaningful sequence such as ‘He bought me a bag’. We can understand that he bought a bag for me, however, if this sentence is placed into different orders - ‘A bag bought me he’ gives it a strange meaning even though it is technically grammatical. The touchscreen task conducted in monkeys as demonstrated in the previous chapters led us to the final phase possible to text one monkey with to ask whether they are able to identify an object by its integrated shape and colour remains unclear. Thus, the combinatorial learning phase would be important to understand the possible evolutionary roots of combinatorial semantics in NHPs.

5.2 Behavioural and Neurobiological Hypotheses

An outstanding scientific question remains unclear is understanding the extent of the combinatorial capabilities of NHPs. I aim to determine whether the macaque monkeys trained with associating speech sounds to visual objects (colours or shapes) are capable to combine these to hear two sounds that determine an object by both its colour and shape by using a novel combinatorial learning task developed in humans (Poletiek et al., 2021). A prior study shows that one of their monkeys can learn a visual combinatorial learning task similar to the paradigm that has been applied to humans (Medam & Fagot, 2016). However, their monkeys overall did not experience stages that are step by step, therefore, it is highly possible that more monkeys (e.g., the ones I have the opportunity to test) have the capability to learn the combinatorial learning task.

Overall, this basis underlies the hypothesis of the touchscreen learning in macaques as follows:

1) that macaque monkeys can learn to identify an object by its shape and colour solely by hearing a short sequence of two sounds that identify those features;

2) that macaque monkeys can do both shape and colour associative learning (mapping a sound to each object feature) but struggle to integrate them, always preferring to identify either colour or shape, or to shift between identifying one or the other but not fully integrating them.

A strong test of this hypothesis is testing the animals with combinations that they would not have experienced during training, as I describe further below to see if they generalise the learning to unexperienced novel combinations.

5.3 Review of Task Process

Below is the design of the combinatorial learning which demonstrates the final touchscreen learning phase in more details. As is shown in the figure (Figure 5.1.), there are several combinations of the colours and shapes (blue cross, blue diamond, orange cross and orange diamond). Additionally, the combinations in the diagonal marked in red were only used as probe trials without juice reward being given to the monkeys. As can be seen from the right side of the figure, to test if the monkey if only using the shape property to identify the object, two stimuli with the same shapes but different colours would be presented on the screen. If the monkey selected the blue cross, then it could be identified that the monkey might just rely on the shape to perform the touchscreen task. Similarly, if the monkey selected the orange diamond, it might suggest that the monkey might just use the colour property when performing the task.

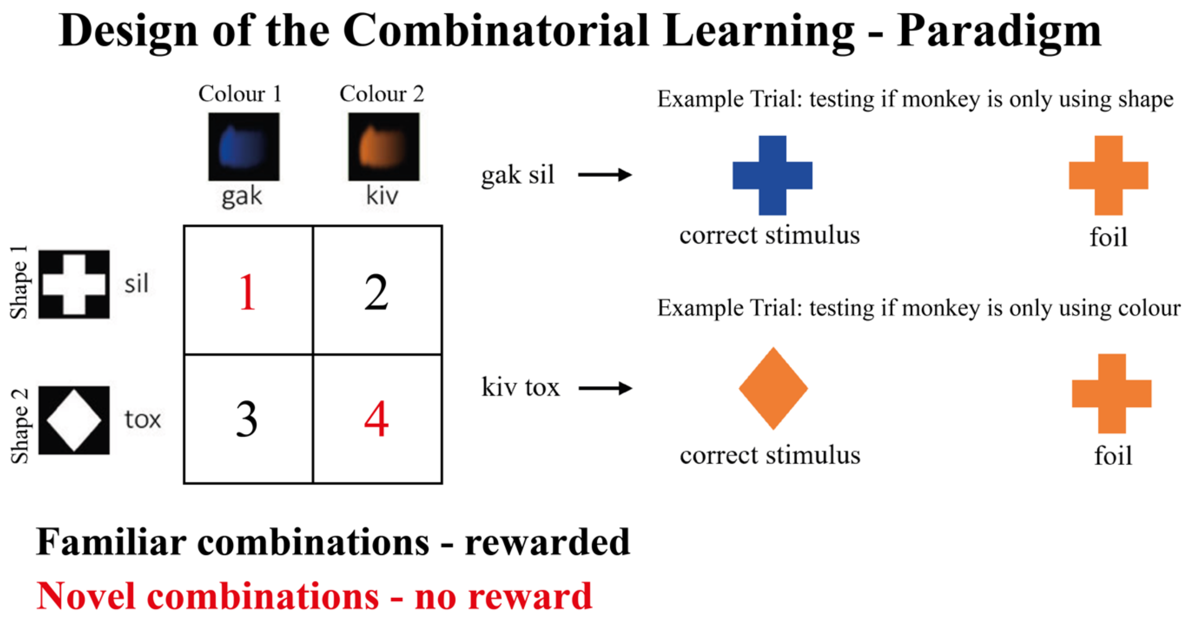


Figure 5.1. Design of the combinatorial learning. The combinatorial learning consists of various stimuli conditions as indicated in the figure above, specifically, there are two combinations for the familiar stimuli (trained combinations) which are orange cross and blue diamond, respectively. Similarly, the novel condition includes the ones which are blue cross and orange diamond. For example, the nonsense words '*gak sil*' is corresponding to a blue cross. To test if the monkey is identifying the object only using the shape property, it can be identified if the monkey chooses the blue cross as is shown in the right panel. This is because these two stimuli share the same shape property while in different colours. Similarly, if the monkey selects the orange diamond, it can be indicated that the monkey is only using the colour property. There are various combinations to test if the monkey is using colour or shape property to identify the object excluding the ones in the diagonal which are 1 and 4 marked in red as these two are the probe trials.

Figure 5.2. illustrates the process of an example combinatorial learning which is similar to what has been explained in the initial behavioural training. By pushing the white spot in the centre of the screen to start the trial, the auditory stimuli are played which are associated with the visual stimuli. For example, the monkey needs to choose between the blue cross and the orange diamond when he hears the sounds '*gak sil*'. If the monkey makes a correct choice, the screen turns green and the juice reward is given to the monkey depending on the number of the consecutive correct trials, the more consecutive correct trials, the more juice reward is given to the monkey. Afterwards, it moves to next trial. Reversely, the screen turns red if the monkey makes an incorrect choice and the auditory stimuli are played again which are associated with the visual stimulus, apart from that, the visual stimulus remains on the screen with the same location to make it clear that this visual stimulus is the one that corresponds to the sounds being

played, and it moves to the next trial. Alternatively, the same trial is repeated and identified as a correction trial, which is used to give the monkey another chance to make a correct choice and re-learn the associations of the auditory and visual stimuli.

Notably, different trial types were presented on the screen in the combinatorial testing phase with customised number of trials. We firstly tried to present the colour and shape only trials on the screen to remind the monkeys the associations between the sounds and visual stimuli. Afterwards, the combination of different colours and shapes were presented. Additionally, we also tried to present the colour and shape trials randomly within a testing session. By doing this, the monkeys could also be reminded of the associations between the auditory and visual stimuli. Eventually, we set a particular number or percent of colour and shape trials at the beginning of testing session. After the completion of these trials, the monkeys progressed to the combinations of colour and shape trials within the testing session. The reward would be given if the monkey made a correct choice in these trial types.

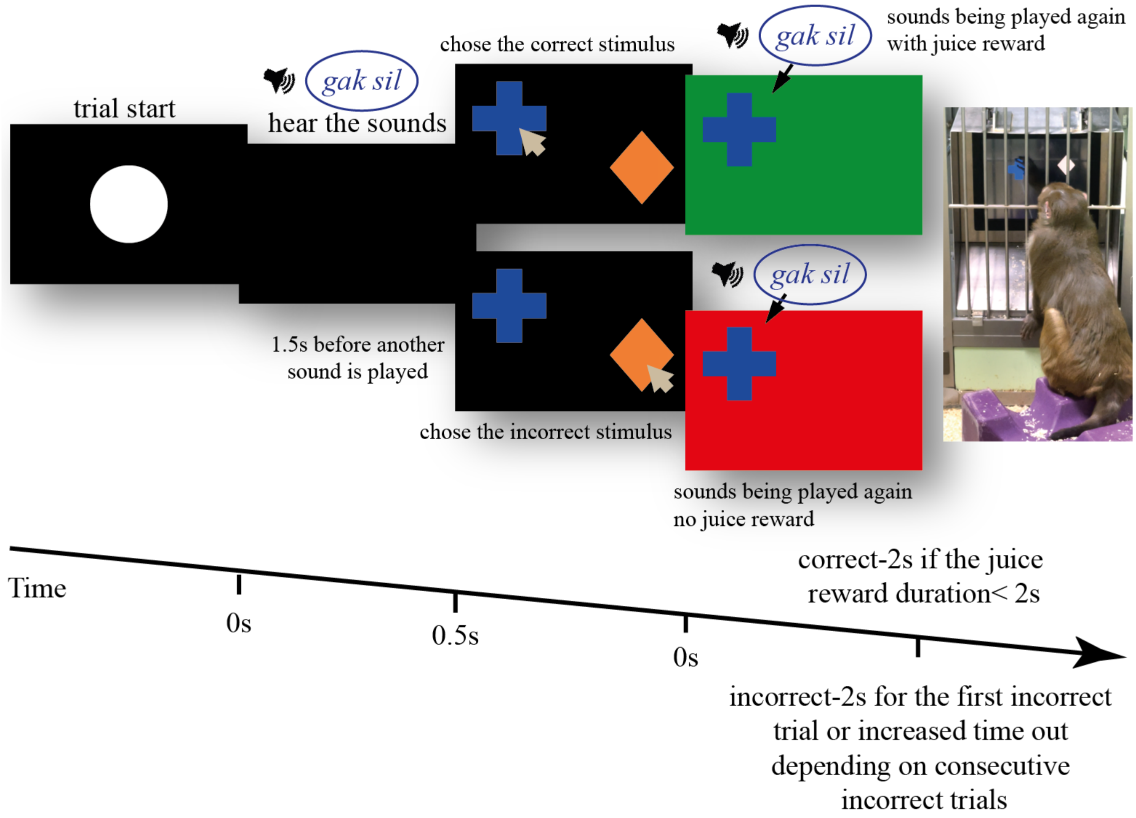


Figure 5.2. An example trial of the combinatorial learning.

5.4 Advancing to Final Phase

During the final phase of the behavioural training among the monkeys, I have tried different ways to improve the monkeys' performance or maintain their performance at high levels, and they are demonstrated as follows. Some of them have been demonstrated before, but more details are given here and particularly focused on the final phase as this was the most important one which indicates whether the monkeys can learn to combine visual stimuli to an object when they hear the sounds associated with the stimuli.

1) Simplify Learning - Reduce the number of shapes and colours from 3 to 2: We decided to reduce the number of stimuli to 2, for example, 2 shapes or colours being presented on the screen during the task as the 3 stimuli might be harder for the monkeys to learn (4 sounds to associate with 2 colours and 2 shapes, rather than 6 sounds to associate with 3 colours and 3 shapes).

2) Improve the reward and delay time schedule: The parameters of the task for reward amounts after N number of correct trials or time outs were modified according to the monkeys' performance to try to improve performance and these were optimised for each monkey as well. Apart from that, we tried to test the monkeys in different times such as mornings or afternoons between 2 and 3 hours or for more limited time on the screen as a reward for working better. Furthermore, we did a further attempt prior to the final phase of data collection to improve the monkeys' performance which is the computer program monitoring the performance during the test, and when performance falls, it gives the monkey a longer penalty time out. More specifically, we were assessing the performance using a sliding window of 5 trials, and if the monkey dips below 60% correct performance, the screen would be shut down for 60 seconds or more. By doing this, we were trying to improve the performance and encourage the monkeys to do more correct trials on the task when they were being tested. It also prevents too many trials below chance, which has been an issue with the touch screen training in the colony, where sometimes the monkeys perform really well and other times lose their motivation or perform at chance. I have also assessed videos of the monkeys' performance to evaluate other factors that may be affecting their performance that I might be able to improve before the final stage of data collection. It was possibly ignored when we tested the monkeys in the colony as they could be affected by the other monkeys such as the

sounds they made or the transport of the other monkeys. Apart from that, it cannot be ignored that the monkeys could not stay in front of the touchscreen all the time during the test as they would not keep drinking all the time.

3) Improved statistical assessment: A permutation significance test to identify strong behavioural sessions was established, based on the number of breaches above chance. It seems to be working because sessions where I change conditions and the monkey falls to chance can be detected, as well as when the monkey comes back up in their performance above chance. This has proven to be a good indication of strong sessions. Also, the behavioural threshold criterion was created to confirm if the monkeys can step into the next testing phase. According to the number of breaches that the monkey breaks through in each session and whether if this is consistent, for example, 3 sessions with above chance breaches per week will determine if the monkey has reached criterion performance and can step into the next phase of training. By improving the task as well as the method of the data analysis, it is promising that we could smoothly test the monkeys with the combinatorial leaning task for the final phase data collection.

4) Improved criteria for assessing above chance performance is as follows: the number of breaches above chance is assessed (relative to permuted chance levels). To monitor the performance over time, we are sliding through each 10 trials, shifting by 1 trial each time in each session. Then we apply a permutation test of the response during these windows shuffling the responses the monkey gives relative to the stimulus conditions for 10 thousand times. Specifically, for each session, a set of random responses is generated 10 thousand times, giving a null distribution of random responses, and from this, where the 5% and 95% confidence intervals are can be calculated. Afterwards, the number of times that the actual performance (based on sliding window analysis) crossed either the 5% or the 95% confidence interval was counted. This helps us to identify strong sessions where the number of breaches above chance are greater than below chance levels. Additionally, as mentioned before, to make sure if the monkey can step into the next phase according to the criterion we have set, the proportion of breaches is calculated based on the timepoints within each session. Specifically, the threshold of the number of breaches is set to 10, and the session is significant if the difference in number of breaches (breaches above 95% CI minus breaches below 5% CI) is higher

than the threshold. If the monkey can do 3 or more than 3 sessions that are significant and consistent within a week, it would be a good indication that it is ready to move to the next testing phase.

5) Advancement of touchscreen device: due to the straight bars of the home-cage, it was not that easy for the monkeys to reach the touchscreen. Considering of the access to the touchscreen, we made some improvements of the apparatus. More specifically, there are two crushes of the home-cage with doors that can be opened without any bars. The touchscreen can be attached to the bottom crush without any limits when the monkeys try to touch the screen. By doing this, the possibility of consistently touching one of the quadrants might be minimised. The advanced touchscreen device was in place in the end, however, due to the limited time of getting the animals involved in the project and the time needed for improving the device, there was not a chance to apply the new device to the monkeys when conducting the testing of the final phase.

By doing all these, the data acquisition of the combinatorial learning phase could be in a good position for the smooth completion of the behavioural project testing in the final phase.

5.5 Final Phase Results with M1 – Familiar Condition

Figure 5.3. is an example session of M1 tested with the combinatorial learning task. This session contains the shape only, colour only and integrated colour and shape trials, meaning that the monkey would go through a few number of colour and shape only trials within the testing session which were randomised. Alternatively, the number of colour and shape only trials could be set to a specific number or percent at the beginning. After the monkey has been tested on the colour and shape only trials, the combined colour and shape objects would be exposed to the monkeys. By doing this, the monkey could have a chance to be remined of the associations of the stimuli before the combined trials started. As can be seen from the figure, the colour only and shape only trials are marked in light blue and pink with stars, and the combined trials are marked in orange. The chance level here is 50% presented in dotted line as the stimuli are 2. Additionally, we are sliding through each 10 trials, shifting by 1 trial each time within each session to monitor the performance over time. There are several trial types being tested in this session including colour only trials, shape only trials and the combined trials, and the percent of each trial type can also be adjusted accordingly. The percent correct keeps at high level

within this session reaching to 95% or more correct, and we can also see many bursts over time which indicates that the monkey could perform very well at some time within a testing session. Notably the performance during Phase 3 was remarkably stronger in terms of percent correct than the Phase 1 and Phase 2 training.

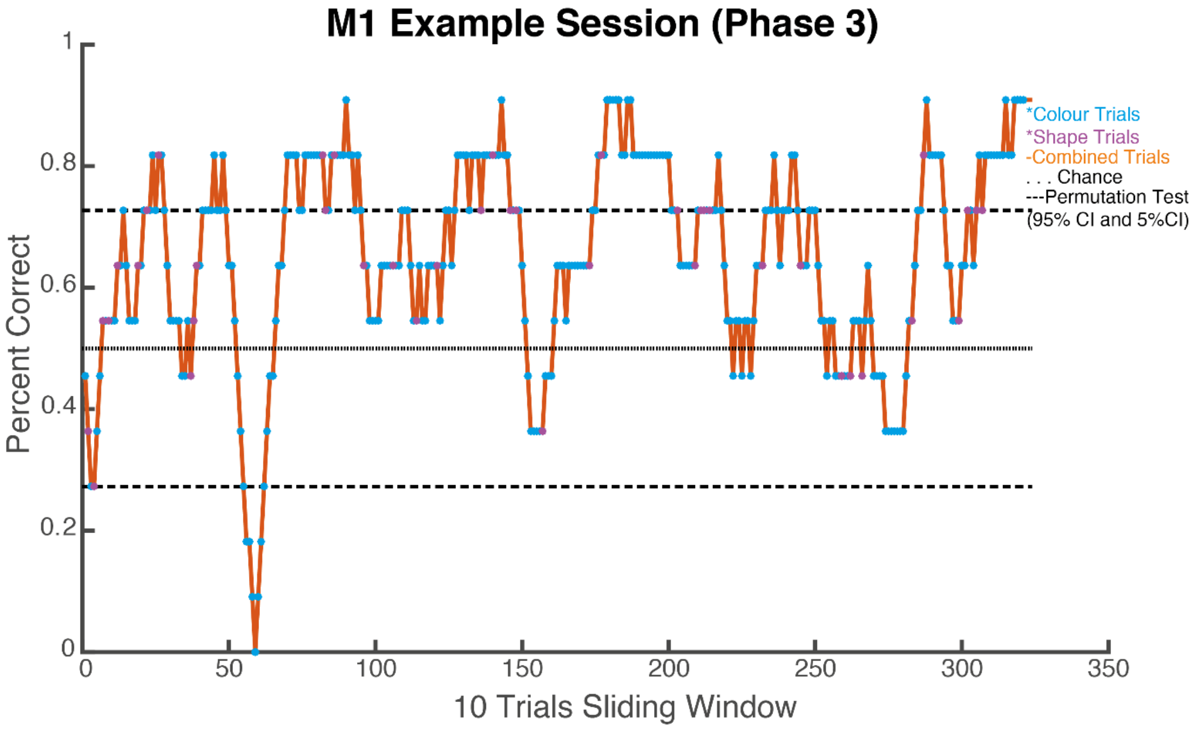


Figure 5.3. An example session in M1 (Phase3). Sliding through each 10 trials, shifting by 1 trial each time in each session. The dotted line indicates the chance performance which is 50% as 2 stimuli are presented on the screen within this testing session, and the dashed line indicates the permutation test. The light blue stars present the colour trials, and the pink stars present the shape trials. The orange line represents the integrated colour and shape trials. Additionally, the percent of these different trial types can be adjusted accordingly. It can be found that there are bursts above the chance level within this session reaching 90% correct. Also, the monkey performs well for each trial types, which means that he can distinguish the colours, shapes, and the combined objects (integrated colours and shapes).

Figure 5.4. shows the distributions of the percent correct for different trial types. To be mentioned, the percent correct of the trials was not a good indication of the monkeys' performance as described in previous chapters as the number of breach differences and the breach trials were used to define whether the monkey performed well on the touchscreen task. However, for Phase 3, the monkey remarkably showed significant increase of the performance when tested with the combinatorial learning task in average. Thus, the average percent correct

became more meaningful for the monkey's performance in the final phase and is the key measure summarised here. As can be seen from the figure, there is a significant difference between these different trial types particularly the colour only, shape only and combined (integrated colour and shape objects) trials, and the percent correct of these trial types are much higher than the others. It can also be found that the percent correct of the combined trials is higher than the shape, colour or opposite foils. Additionally, the ANOVA test conducted with the comparison of combined trials to shape foils, colour foils and opposite foils shows significant differences (F(3,192) = 78.86, $p < 0.001$), indicating that the monkey can learn to identify an object by its colour and shape property.

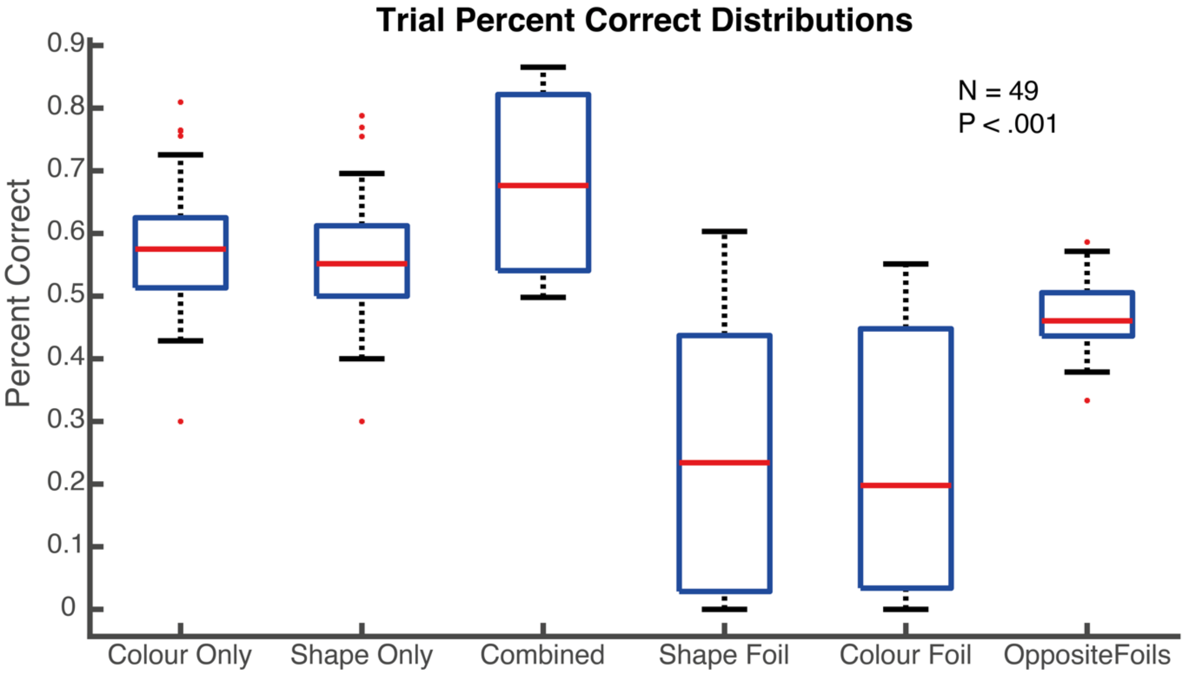


Figure 5.4. Familiar combinations of shape and colour trials with reward. 21 sessions with familiar trials only and 28 sessions with familiar and novel trials. Only familiar trials are analysed and presented in this figure; One-way ANOVA for multiple comparisons; combined > foil performance, N = 49, $p < 0.001$, Factor degrees of freedom: 3, Error Degrees of Freedom: 192; One-way ANOVA for multiple comparisons; combined > all other trial types performance, N = 49, $p < 0.001$, Factor degrees of freedom: 5, Error Degrees of Freedom: 288 (error bars are standard deviation, colour only: colour trials; shape only: shape trials; combined: integrated colour and shape trials; shape foil: same colour with different shapes; colour foil: same shape with different colours; opposite foils: distinguished shapes with different colours).

To investigate whether the combined trials were the best among all trial types in the familiar condition, I also conducted the ANOVA to compare the performance of combined trials to all the other trial types. This test showed a significant difference (F(5,288) = 75.49, $p < 0.001$), indicating that the monkey performed the best in combined trials.

Additionally, the percent correct of different trial types over sessions in Figure 5.5. shows whether the monkey showed increasingly learning across sessions in the final phase. It can be seen from the figure that the monkey's performance was increased over time (sessions) in shape only, colour only and combined trials. On top of that, the monkey performed the best in the combined trials, which indicates that the monkey can learn to identify the object by its colour and shape features.

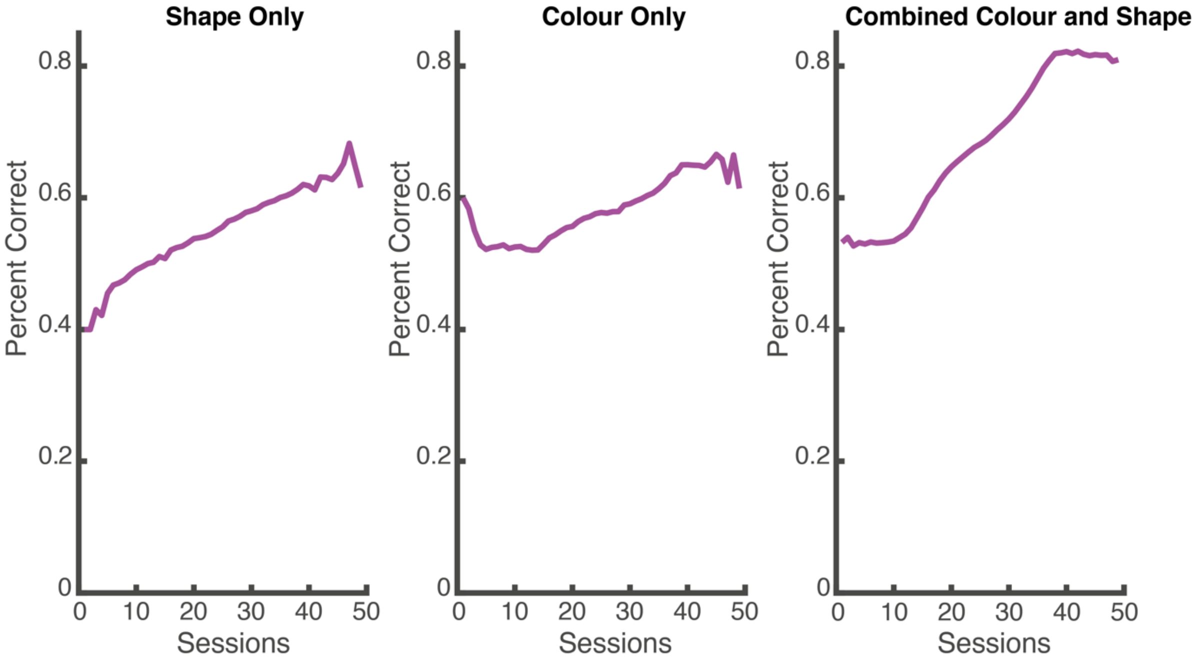


Figure 5.5. Learning curves of Phase 3 – familiar condition. 49 sessions in total including colour or shape only trials as well as integrated colour and shape trials (shape only: shape trials; colour only: colour trials; combined colour and shape: integration of colour and shape trials).

Moreover, Figure 5.6. presents the error rate of different foils in the familiar condition including shape foils, colour foils and opposite foils. These are errors (incorrect trials) that decrease over the testing sessions, indicating that the monkey more readily was able to reject the incorrect stimulus in both shape and colour foil conditions. Errors on opposite foils remain flat. This suggests that the monkey improved the ability to distinguish the shape or colour properties from the foils (compare Fig.5.6. errors to Fig.5.5. percent correct). Also, it showed that the monkey could make correct choice without relying only on one of the features, although the monkey continued to make errors for the opposite foils.

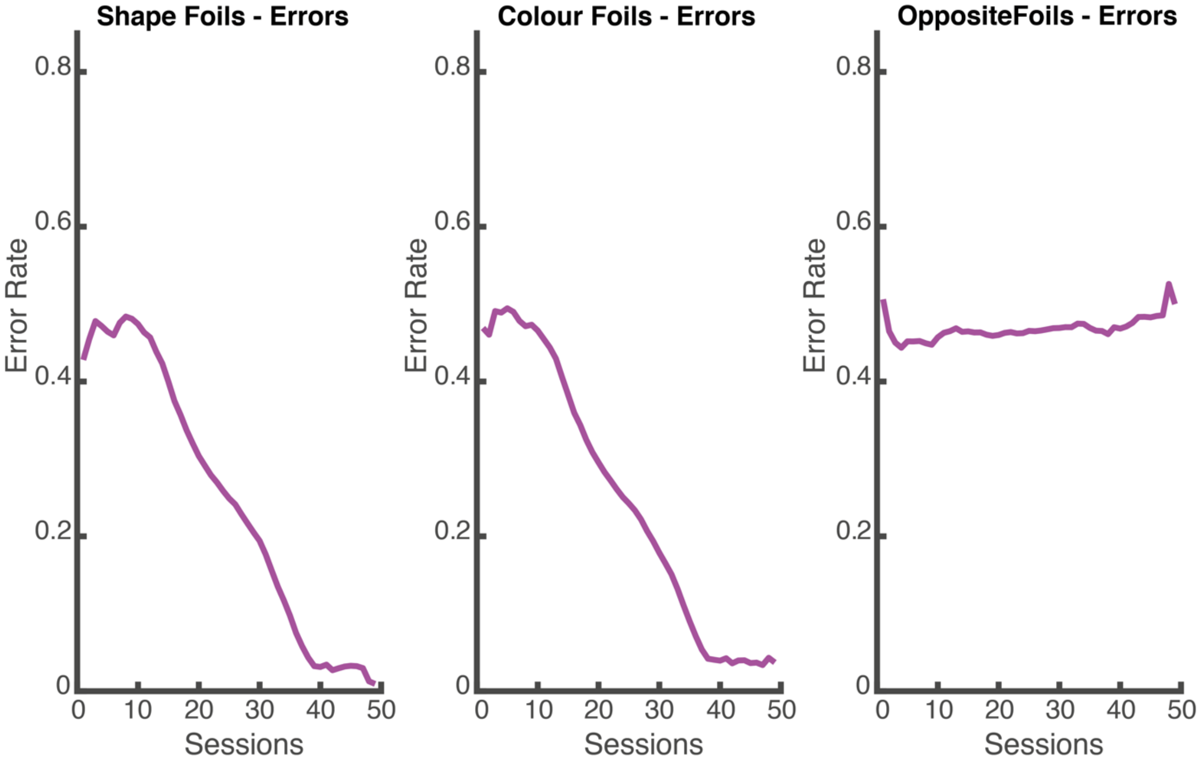


Figure 5.6. Error Rate of Phase 3 – familiar condition - foils. 49 sessions in total including colour or shape foils as well as integrated colour and shape foils (shape foils: foils with same colour but different shapes; colour foils: foils with same shape but different colours; opposite. foils: foils with distinguished integration of colours and shapes).

To summarise, the results suggest that the monkey can learn an object by its colour and shape property, and the evidence of the increasingly learning is proved across sessions. However, it would be important to see whether the monkey can show generalisation to novel trials.

5.6 Final Phase Results with M1 – Novel Condition

Regarding to the novel conditions, these were the conditions which the stimuli were not previously presented together, and they were tested here as probe trials with 25% of the trials in one testing session. Specifically, the stimuli in the novel conditions were the ones with novel combinations of the speech sounds that referred to a visual object. For example, these two speech sounds '*gak sil*' referring to a blue cross, and the monkey would hear the sounds one after another with an interval of 1.5 seconds in between before the visual stimuli was presented. In the meantime, the foil stimulus would be presented on the screen, and the monkey needed to select from the novel visual stimulus and the foil. However, there would be no juice reward given to the monkey no matter the monkey made a correct choice on the novel visual stimuli or not. By doing this, the possibility of the association between the reward and the novel stimuli could be avoided.

Figure 5.7. shows the distributions of the percent correct for different trial types in the novel condition. As is shown in the figure, there is a significant decrease of the percent correct in the novel condition. Specifically, the percent correct of colour and shape only trials stays stable compared to familiar condition. Apart from that, the result shows the lower percent correct of the novel combined trials which is not even reaching to the chance level (50%). Also, the percent correct of novel shape foils as well as novel colour foils is lower when compared to familiar condition and is much lower than the novel combined trials. On top of that, when tested with the ANOVA test with the comparison of novel combined trials to novel shape foils, novel colour foils and novel opposite foils, the result shows significant differences ($F(3,108) = 71.79$, $p < 0.005$), indicating that the monkey performed very differently from its performance on the rewarded familiar trials above. Namely, when the novel trials were unrewarded and probed (25% of the trials that the monkey experienced), the monkey's strategy seemed to change entirely as if it were treating the novel stimuli as completely different from the familiar ones. For example, the performance although at the highest level in Figure 5.4. for the combined shapes rather than the colour only or shape only trials. In this case, performance to the combined shape stimuli is well below chance, as is the performance for many of the foil stimuli. Therefore, it is difficult to conclude whether the monkey could generalise its performance to novel stimulus combinations given the drastic change in performance with the probed unrewarded novel combinations. Note that the same stimulus conditions were used here, they were just presented in novel combinations not previously presented together or rewarded.

We tried to minimise the potential impact on the monkey's behaviour, however, the proportion of novel 'probe trials' which were unrewarded is fairly high at 25%, meaning that the monkey could regularly observe a trial that was unrewarded 'probe trial' and thereby alter its response to these sets of trials.

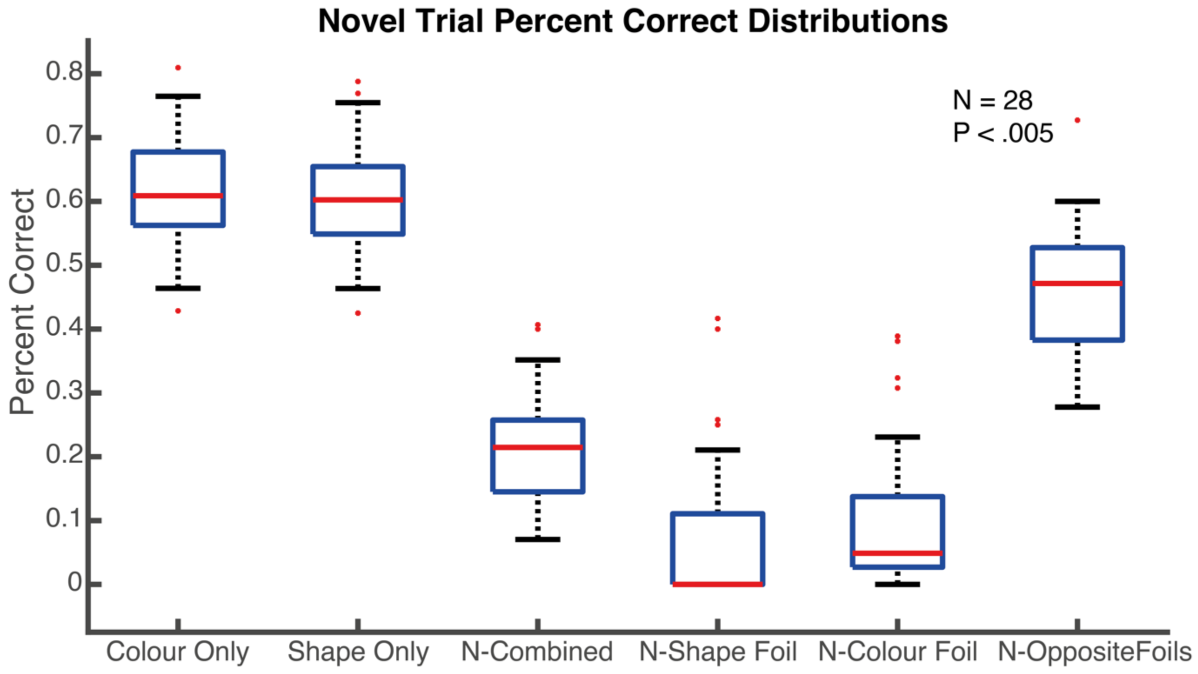


Figure 5.7. Novel condition – unrewarded. 28 sessions with familiar and novel trials included only novel trials are presented here, N = 28, p < 0.005, Factor degrees of freedom: 3, Error Degrees of Freedom: 108. The novel combinations are probe trials without juice reward during the testing sessions. It is shown in the figure that the monkey did not perform well among these sessions and the percent correct are below the chance in each of the different trials types (error bars are standard deviation, colour only: colour trials; shape only: shape trials; N-Combined: novel combinations of the shapes and colours that the monkey has never been trained; N-Shape foil: novel stimulus with the same colour but different shapes; N-Colour foil: novel stimulus with the same shape but different colours; N-Opposite Foils: novel stimulus with different combination of shapes and colours trials).

In comparison with the familiar condition, the percent of shape and colour only trials in novel condition keeps increasing across sessions as is shown in the figure (Figure 5.8.). However, it shows an opposite profile in the novel combined trials, which decreases from the start of the testing to the end. This is interesting in that the monkey was still able to perform well on the shape and colour only trials, however, the monkey showed a gradual decrease in the learning of novel combined trials, which is unlikely a decrease in learning but more likely a change in its response strategy to the ‘probe trials’.

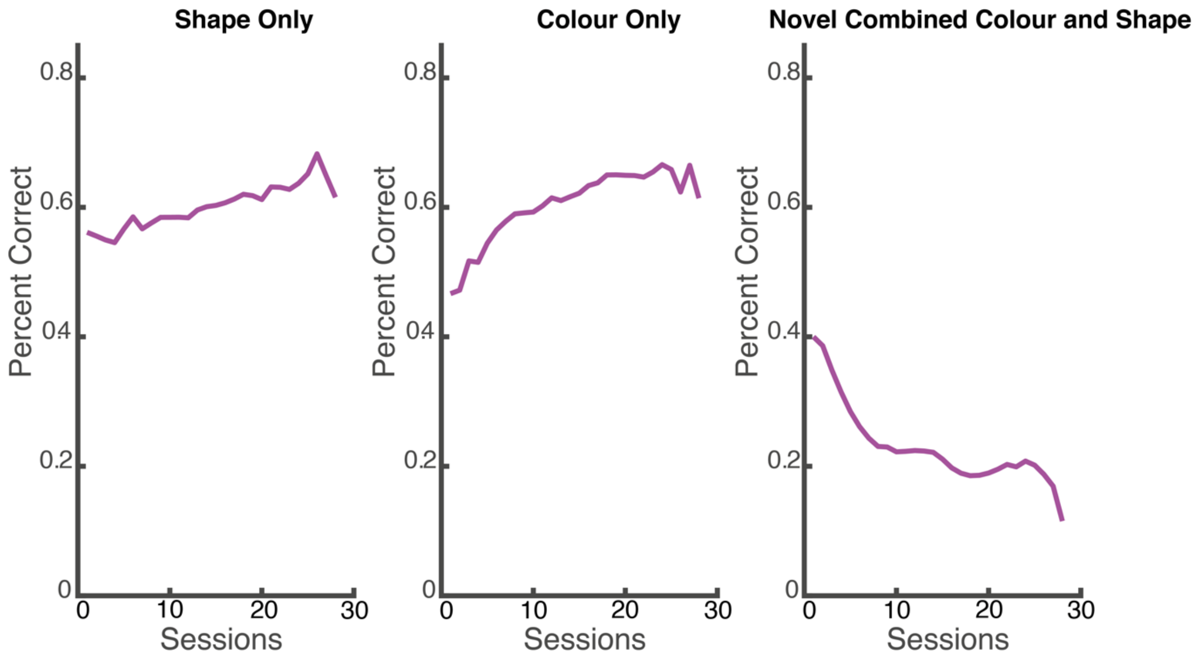


Figure 5.8. Learning curves of Phase 3 – novel condition. 28 sessions in total including colour or shape only trials as well as novel integration of colour and shape trials (shape only: shape trials; colour only: colour trials; novel combined colour and shape: novel integration of colour and shape trials).

Additionally, Figure 5.9. shows the learning curve of different foils in the novel condition. as can be seen from the figure, there is a decrease across sessions over time in both the novel shape foils and the novel colour foils. Furthermore, the novel opposite foils stays stable which is similar to the one in familiar condition. The decrease in novel shape foils and the novel colour foils is consistent to the result in the familiar condition, suggesting that the monkey was able to learn and distinguish the shape and colour properties in either familiar or novel conditions, however, not showing the generalisation to novel combined trials.

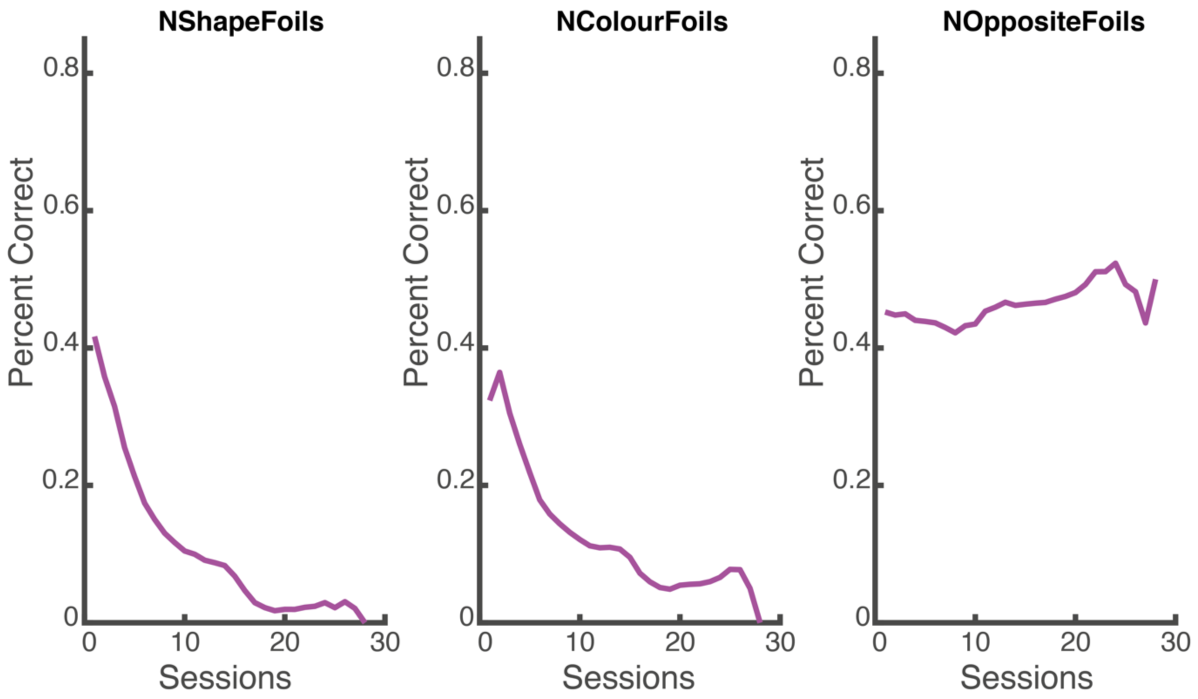


Figure 5.9. Learning curves of Phase 3 – novel condition - foils. 28 sessions in total including novel colour or shape only trials as well as novel integrated colour and shape trials (NShape foils: foils with same colour but different shapes; NColour foils: foils with same shape but different colours; opposite. NOpposite foils: foils with distinguished integration of colours and shapes).

5.7 Discussion

In the combinatorial learning phase, we aim to investigate whether NHPs are capable of identifying an object by its joint colour and shape property. Before we started the final testing phase, we have tried many ways to improve the design of the touchscreen task for increasing the monkey's performance such as reducing the number of stimuli from 3 to 2 to decrease the difficulty of the touchscreen task for advancing to the combinatorial learning phase. Additionally, we have also tried to balance any bias towards the shape or colour properties to minimise the possibility of that the monkey relies on either the shape or colour to select among the visual stimuli. The key finding was that the monkey could learn to identify an object by its colour and shape in the familiar trials that were rewarded. In other words, with reinforcement learning the monkey showed remarkably strong performance (stronger percent correct than the initial testing phase) highest for the two sounds that identified the combined object by correct shape and colour, rejecting the foils that differed on one or both dimensions.

However, the generalisation to novel combined shapes and colours, which were tested with unrewarded probe trials, was not found in the monkey. This is similar to what Medam and Fagot have found in their study of the baboons as one out of seven baboons involved in their first experiment successfully distinguished objects with compound labels (Medam & Fagot, 2016). It is worth noting that in that study there did not appear to be the equivalent of our novel testing combinations, although the authors note that performance was high for the combinations in the monkey during the initial testing phase when the stimuli were technically not yet familiar. Additionally, all baboons tested in their second experiment all failed to interpret novel combinations of symbols learned individually (Medam & Fagot, 2016). This might suggest that at least two NHPs (baboons in their study, one macaque in ours) might be able to learn to identify an object by integrated properties, which altogether is an interesting observation. We do not know if M2 in this study would have shown similar results because we had to complete with that monkey before it could be tested in the final phase. Nonetheless, even if there is considerable variability in whether most baboons/macaques can learn to map visual to visual or auditory to visual stimuli that combine features, at least two individuals can learn what may be the most language-like task that has been tested to date in NHPs.

Many studies have shown that NHPs and children can perform the touchscreen task and learn the associations between the auditory and visual stimuli or the cue (label) referring to the visual object (Calapai et al., 2022; Ghasemian et al., 2021b; Medam & Fagot, 2016; Vonk & Rastogi, 2019), however, not many of them have used the artificial sounds that referring to a visual object, and to investigate the combinatorial learning abilities whereby the monkey hears two sounds that refer to different visual properties that need to be integrated to identify an object by its joint properties. The touchscreen task we designed might not be that easy for the monkeys as they only need to learn the associations between the auditory and visual stimuli, which are known to be difficult for monkeys to achieve (Rinne et al., 2017), but also they need to learn the combinations of the colour and shapes of an object. This is also the reason why we reduced the stimuli from 3 to 2 as what has been mentioned before because it was difficult for the monkeys to perform well with 6 different sounds identifying 3 shapes and 3 colours.

Additionally, we tried to progress the monkeys through different testing phases step by step, and to avoid the possibility that the monkeys are relying on either the shape or colour property. This makes the touchscreen more complex and time-consuming as well. However, the results suggest that the monkey did not rely on the shape or colour property to select the visual objects when tested with the final phase, meaning that the monkey can learn to distinguish different features of the visual object. Moreover, the reason for including the colour and shape only trials within a testing session in the final phase was that to remind the monkey of the associations between the stimuli, and the results have proved that this worked very well as the performance increased over contiguous sessions in the shape only trials, colour only trials and the combined trials.

Apart from that, one of the possibilities that the monkey did not show generalisation to novel combinations might be because of the limited testing sessions and the limited novel foils that can be used in the final phase. Another possibility is that I knew I had limited time to test the macaque so although we would have preferred to have fewer probe trials (less than 10%) so that the monkey did not change its response strategy by noticing the novel stimulus conditions, we opted for 25% probe trials which seemed to have changed the monkeys behavioural response strategy (performing well below chance on many of the novel tested conditions). Therefore, we do not know if the monkey could possibly show generalisation to novel combinations if some extra time was allowed, and fewer probe trials were used. We also could have rewarded half of the novel trials at random rather than not rewarding any of the probe trials. It is also possible that the monkey could have been confused while tested with the novel combinations as some of the visual stimuli were used as the foils in the familiar condition even though they were not associated with any of the auditory stimuli. Therefore, it could possibly have some impact on the monkey’s behavioural strategy when they were trying to select among the visual stimuli as the monkey had seen the visual objects before but without any sounds.

It is possible that the monkey just learned a fixed association between a longer sound (two sounds in a row) and a richer image with two attributes (colour and shape) compared with the previous phases (colour/shape only). This account finds support by the improved performance across time (Fig.5.5.) as the learning curve showed that the monkey’s performance kept increasing, while errors to the foils generally decreased. However, the performance of colour/shape only trials in the combinatorial learning phase also kept increasing no matter in familiar or novel conditions (Fig.5.5 & Fig.5.8.). This might suggest that the monkey did not only learn the combined colour and shape trials with longer sounds and richer properties, but also learned the colour or shape trials separately with just the colour or shape property. This could be further explained by the fact that we introduced the colour/shape only trials in the combinatorial testing phase with a fairly high percentage of these trial types to remind the monkey about the associations between the sounds and colour/shape properties. And it was found that the monkey showed increased performance among all the trial types in the familiar condition. This suggests that introducing the colour/shape trials within the testing session in the combinatorial learning phase could help the monkey to firstly associate the sound and colour/shape separately and then integrate the colour and shape properties into an object. Therefore, it is more likely that the monkey can integrate the colour and shape into an object by hearing the associated sounds rather than identifying a coloured object with a longer fixed sound. This strategy is unlikely to help the monkey when different combinations of sounds were presented, unless the monkey memorised all the combinations, which we cannot rule out. Therefore, although it seems unlikely that the monkeys conducted the task by memorising all the sequence sound pairs as one unifying ‘complex sound’ and linking these to the specific object with the colour and shape properties, we cannot rule out the strategy without having tested many more combinations.

To conclude, our study progressed the monkeys through the testing phases step by step, allowing them to learn the touchscreen task in a smooth way. We have also tried different methods to improve the touchscreen task and to improve the monkeys’ performance. This allowed us to test M1 on the final phase with remarkable results on the high levels of performance for the combined stimulus conditions, and the monkey rejected the wrong foils. The results indicate that the monkey can learn to identify the object by its colour and shape properties, but not show generalisation to novel combinations, which if I had more time, I would conduct in different ways to obtain stronger evidence that the monkey could or could not generalise its combinatorial learning to novel combinations.

Chapter 6: Comparative fMRI of Audio-Visual Sequence Processing in Macaque Brain

Parts of this chapter have been presented as

Zhang. Z., Wilson. B., Muers. R., Milne. A., Kikuchi. Y & Petkov. C. (2022) Complexity of Primate Cognition and Neural Bases of Sensory Convergence. North East Postgraduate Conference, Newcastle, 2022.

6.1 Introduction

Language abilities and complex mental constructions in humans allow individuals to identify the structure of sensory sequences in any sensory modality. Although monkeys do not have language, their cognitive system is possibly much more sophisticated than currently thought (Ferrigno et al., 2020). For better understanding the evolutionarily conserved neurobiological specializations for language (Wilson et al., 2015), a comparative human and monkey fMRI task was designed to help us to understand audio-visual sequence processing in the primate brain. Specifically, revealing how the brain integrates sensory inputs.

The primate brain is highly multisensory. For example, there are auditory inputs into visual cortex and vice versa, and in fact some have argued that there is no ‘unisensory’ cortex anywhere in the brain (Ghazanfar & Schroeder, 2006). However, this might apply for basic auditory/visual properties (like when a sound/object needs to be identified very quickly). Frost et al. (Figure 6.1.), however, argue that when it comes to integrating information over time (sequential information) the sensory systems are surprisingly separate (Frost et al., 2015). The model they schemed suggests that sequence learning does not occur across modalities, with separate brain regions processing the auditory or visual input solely, which accounts for stimulus and modality specificity.

This makes a very interesting prediction about how the primate brain processes auditory and visual sequences that have the same structure over time. This forms the basis of the hypothesis as follows.

6.2 Hypothesis

Some parts of the superior temporal sulcus (STS) and the prefrontal cortex will process the structure of the sequence regardless of whether they are auditory or visual, showing very similar fMRI patterns to the auditory and visual sequences obtained in separate experiments and testing sessions.

Alternatively, different patterns of activity may occur for auditory and visual sequences, meaning that these areas in the brain are modality specific. This would be consistent to the model Frost et al. proposed that the sensory system is sperate when it comes to integrating the auditory and visual sequential information over time.

To test how the primate brain processes auditory and visual sequences that have the same structure over time, we have designed an experiment to conduct with human and macaque fMRI. Two prior students started the data collection but were not able to complete it and I would like to complete this project and be a co-leader with them (Alice Milne - collected the majority of the human fMRI data; Ross Muers - collected the data in Monkey 1 and the remaining human fMRI data; I have collected the Monkey 2 data and completed the pre-process of the datasets, and will be involved in further analysis in the future). The three of us could be joint first authors since we all contributed to this extensive project to ensure its completion. Below is a summary of the data that has been collected for the fMRI project.

Table 6.1. The number of sessions of fMRI in monkeys and humans

Number of sessions	Monkey 1	Monkey 2	Human Participants
Auditory sessions	19	26	15
Visual sessions	16	31	15

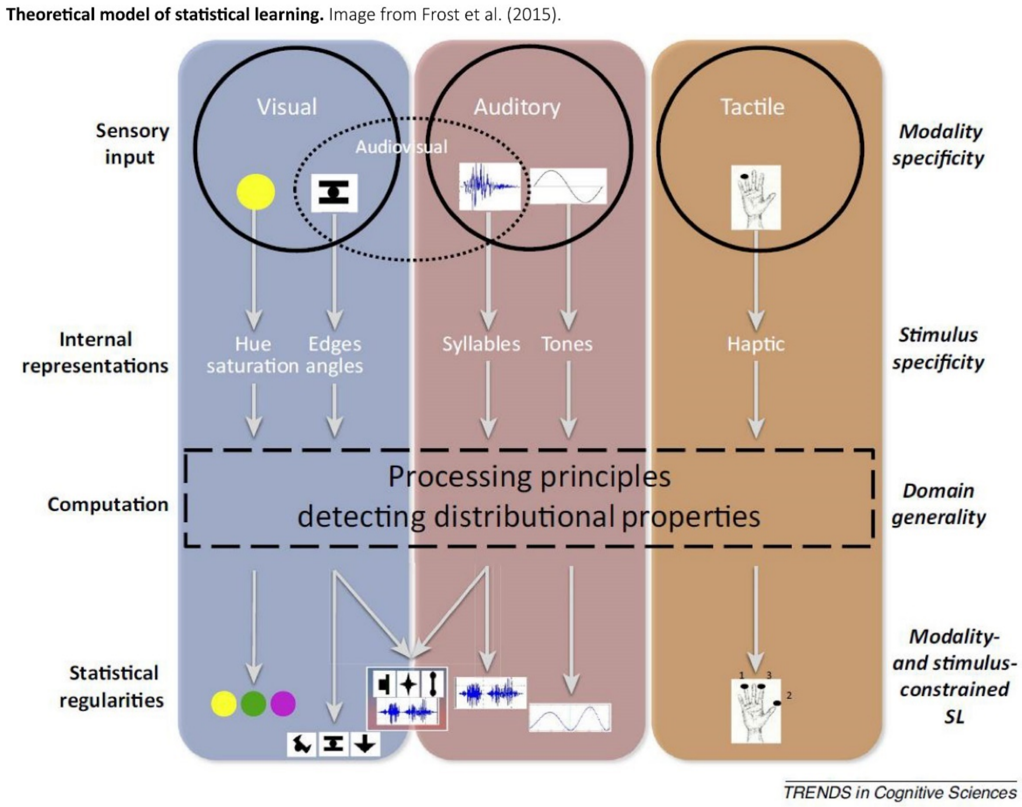


Figure 6.1. Theoretical model of statistical learning. Image from Frost et al. (2015)

6.3 Subjects

The fMRI data was collected in a female rhesus macaque (*Macaca mulatta*) with a Bruker 4.7 Tesla scanner (Bruker BioSpin, Etlingen, Germany) being placed in a vertical way in the scanning room, and the monkey was housed in the Nonhuman Primate (NHP) area of the Comparative Biology Centre (CBC), Medical School, Newcastle University, the UK under Home Office licensing and regulation (Animals (Scientific Procedures) Act 1986), including project license and personal licenses for all the personnel working with the primates. The female (age: 6, weight: 6 kg, in-pair with another male) was acclimatised to head immobilisation before being taken out to the scanner and was trained on a visual fixation task (in a square chair rather than the scanning chair in the lab) to ensure the high precision eye tracking data.

6.4 Stimuli

The auditory stimuli were computer generated sound effects and sampled at 22050Hz, and each sound was 410ms long. The sounds were combined into sequences with a 150ms inter-stimulus interval and 2650ms of sequence duration. The visual stimuli (9cm x 9cm) were generated using Adobe Photoshop (version 13.0.1.2) with a duration of 410ms, and presented on a grey background with a 150ms inter stimulus interval. Psychtoolbox 3 (Psychophysics Toolbox Version 3) was installed in Matlab (Matlab R2017a) to be applied in both the auditory and visual experiments.

6.5 Methods and Procedures

To test this, we have created an experiment (Figure 6.2. and Figure 6.3.) as follows, with the auditory and visual scanning conducted separately and using a multi-variate pattern analysis (Representational Similarity Analysis) to test whether the auditory/visual representations of the sequences as similar in certain parts of the primate brain or remain dissimilar in many or all of them. All the auditory stimuli were computer generated sound effects, and all the visual stimuli were arbitrary monochrome objects. The experiment is based on a behavioural study (Milne et al., 2018) which showed that monkey and human behaviour is sensitive to the sequencing dependencies (transitional probabilities) and similar across the auditory and visual modalities. This study provided the behavioural basis for the fMRI study.

6.5.1 Eye-Tracking Method and Analysis

The eye-tracking paradigm was used to record the looking response by macaque monkeys throughout the experiment (infra-red eye tracker, Arrington Research), which was previously developed by Wilson et al. and implemented as needed in the fMRI experiments. Monkeys were initially trained to look at the yellow spot in the centre of the screen for a short period of 3 seconds, the fixation spot disappeared from the screen and the juice reward was given to the monkey when successfully maintained the fixation period. Otherwise, the trial aborted with no juice reward if the monkey broke the fixation. After the fixation period, it continued to the stimulus presentation period, that said, the auditory or visual stimuli would be presented on the screen sequentially and the monkey was able to freely move the eyes as the stimulus could be played from either the left or right side of the headphones in audio modality and be presented from the left or right side of the scree in visual modality. The eye movement was tracked in both the fixation training and the MRI scans. There was no delay after the trial aborted and the next trial started.

Each of the audio/visual trial lasts about 10 seconds (audio-10.2 seconds, visual-10 seconds). The eye-tracking data analysis contained a 3-second baseline in each of the trial as the monkey was required to fixate on the spot for a duration of 3 seconds, and the subsequent 6 seconds during which period the audio/visual sequences were presented. To calculate the duration of looking responses in the auditory experiment, I can initially calculate the baseline variability in the eye movement during the final 2 seconds fixation period of each trial, and the initial 0.5 second during which the spot requisition on can be excluded as the monkey saccades to the spot. Looking responses to the test sequences were defined as looks towards the presenting audio speaker from either the left or right exceeding 3 standard deviations of the variability in the baseline fixation period (Wilson et al., 2013). In the visual experiment, looking responses to the visual sequence were recorded when the gaze fell within the window (Fig.6.2, an analytical inclusion window) which can be adjusted in terms of the size. It was defined within the range of ±10º elevation and 12-35º azimuth around the visual stimuli location.

6.5.2 Exposure Phase

Both the exposure and testing phase were conducted in the scanner, and there were no scans during the exposure phase but with the juice reward given to the monkey to keep the motivation of the animal. The monkey was presented with sequences that follow the rules of the Artificial

Grammar (AG) as is shown in the figure (Fig 6,2.), and it was required to listen to or look at the sequences as a response during the exposure phase. these 8 exposure sequences were presented randomly with no specific order repeating 6 times in about 5 minutes. Also, the juice reward with 500 milliseconds was given to the monkey randomly to maintain the motivation. During the exposure phase, the stimuli were played from both sides of the headphones in the auditory experiment and the stimuli were presented in the centre of the computer monitor in the visual experiment.

6.5.3 Testing Phase

Similarly, the monkey was presented with the audio/visual sequencing stimuli during the testing phase. The monkey was tested over multiple sessions including separate auditory and visual sessions. Specifically, 6 runs were conducted with 38 trials in each of the audio/visual sessions. However, there might be slight differences in some of the sessions such as the number of runs depending on the performance of the monkey or there could be some situations that the scanner did not work properly. In both auditory and visual experiments, each trial started with a white spot in the centre of the screen, the monkey was required to fixate on the spot for 3 seconds before the stimuli being played (auditory) or presented (visual).

In the auditory experiment, the speakers (in the lab for necessary training) or headphones (in scanner) were put on the monkey and the auditory sequences stimuli were played through the headphones in the auditory experiment. The sounds were presented at about 75dB SPL (A weighting; calibration: XL2 sound level meter, NTI Audio). The sequences that matched or violated the Artificial Grammar sequencing rules were tested.

In the visual experiment, the stimuli were visual, but the Artificial Grammar sequencing rules were the same. The stimuli were presented sequentially from either the left or the right side of the screen after the fixation period in random orders. The monkey was able to move the eyes freely during this period as the stimulus was presented one after another in a sequence.

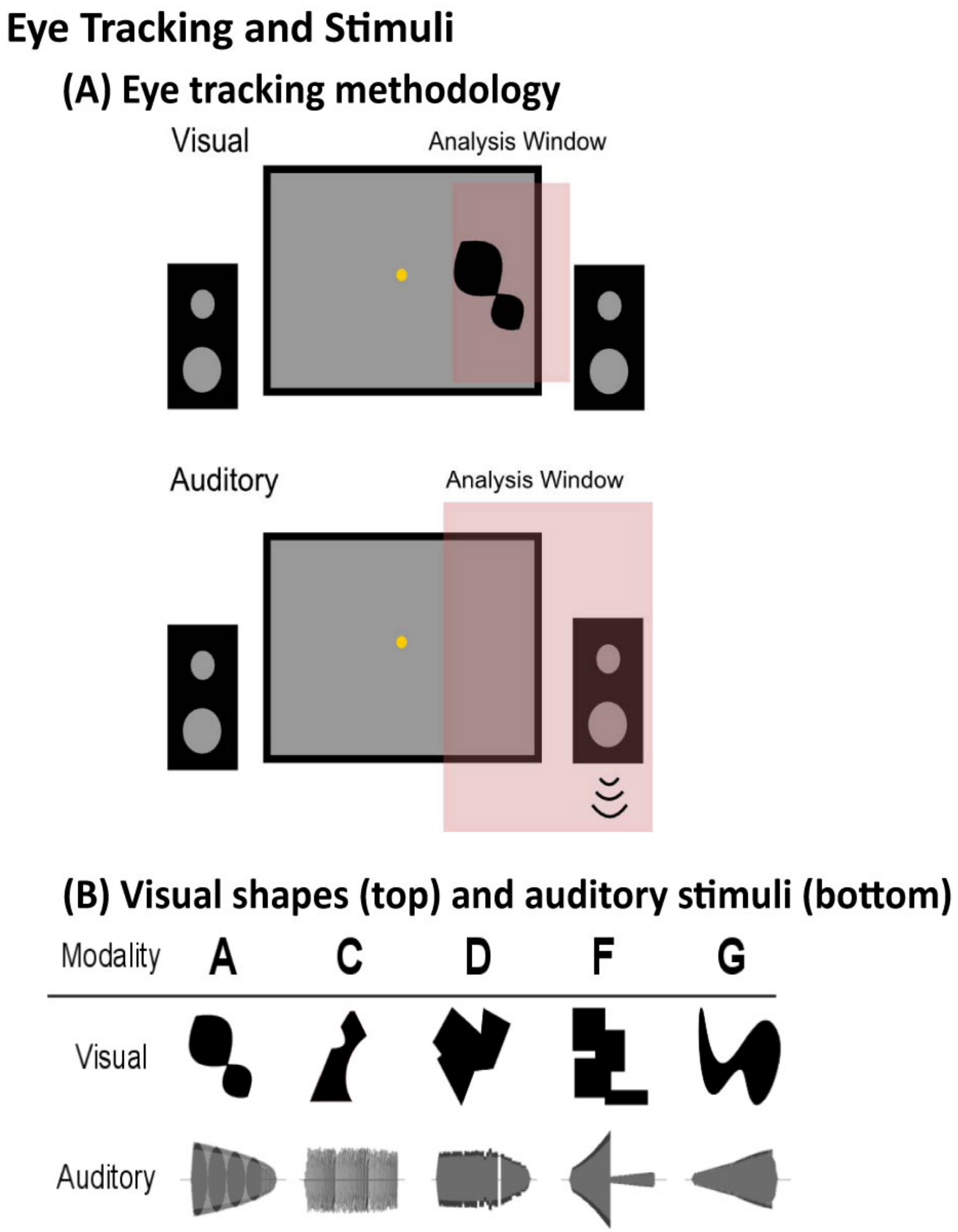
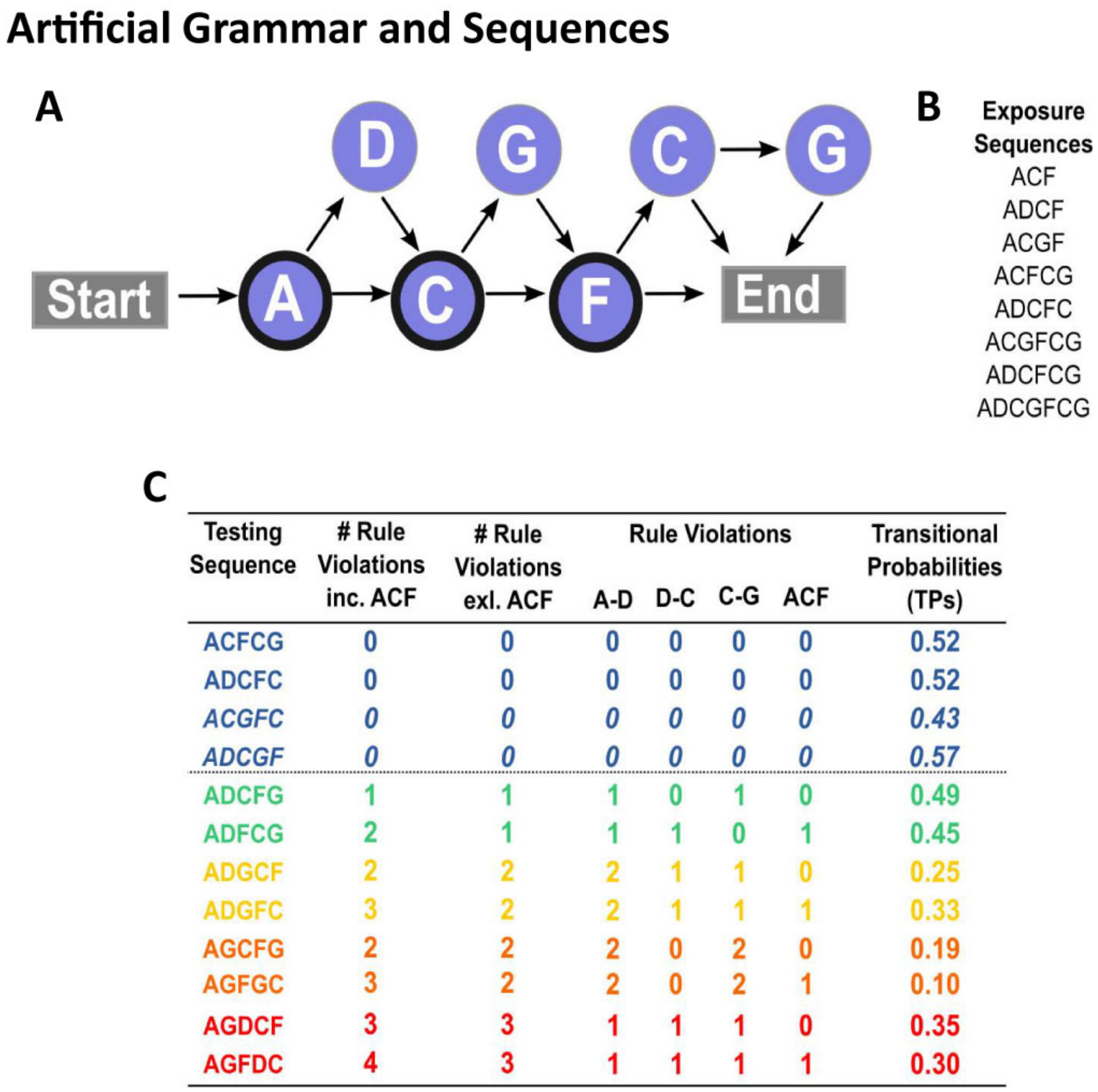


Fig 6.2. Eye tracking and stimuli. (A) Eye tracking methodology. In the auditory experiment, the analysis window contains all responses that exceed the eye movement variability by 3 standard deviations (SD) during the previous baseline fixation while in the visual experiment, the analysis window overlaps with the stimulus position. (B) Visual shapes (top) and auditory stimuli (bottom). Image sourced from Alice Milne’s PhD thesis.

The Artificial Grammar (AG) is based on several rules giving the order that items will appear in a sequence. All sequences are structured according to the AG state transition diagram shown in Figure 6.3A. Legal sequences are generated by following any of the arrows from the start to the end. For instance, the longest sequence can be generated from the AV is **ADCGFCG** while during the process, **A** can go to **C** or **D**, and **C** can go to **F** and **G**. Besides, the elements circled in black in Figure 6.3A are three compulsory elements meaning that they must appear in each of the sequences being generated, and they can be adjacent to one another, meanwhile, the other elements can appear in the sequence to separate the compulsory elements or create non-adjacent relationships. The behavioural study and the stimulus sequences were created to allow analysis of transitional probabilities, specific rule violations, the number of rule breaks and whether the adjacent or non-adjacent relationship is violated. These 'behavioural models' will be used in the fMRI analysis.

The exposure sequences are shown in Figure 6.3B generated from AG as well. They are used to allow the subject to become familiar with the structure of the sequences and to induce what is called 'statistical learning'. These exposure sequences vary in length from the shortest 3 elements (**ACF**) to the longest 7 elements (**ADCGFCG**), but the testing sequences are all matched in length.

Apart from that, all test sequences begin with A and contains 5 elements, consisting of 4 consistent sequences (in blue in Figure 6.3C). Specifically, 2 of these sequences are presented during the exposure phase and the other two are novel that do not appear in the exposure phase. Importantly, to avoid the subjects from rote memorising the exposure sequences, novel consistent sequences are included in the testing set. The violation sequences are created according to different aspects of the AG below the dotted line in Figure 6.3C referred to as 'rule violations'. Specifically, three rules are established here which are adjacent: 1) the sequence must proceed by A if D is not presented; 2) If A is proceeded, it must follow by C; 3) it must be proceeded by C if G is not presented. Rule 4 is non-adjacent which must include A, C or F, but it is not necessarily needed to be presented in an adjacent sequential order. The violation sequences consist of different numbers of rule violations including adjacent and non-adjacent sequences as well as varying mean transitional probabilities (TP). It calculated the amount of times a transition occurs during exposure relative to the frequency of the element present on its own, indicating the possibility that one of the elements will be followed by the next with relation to the frequency.



Testing Sequence	# Rule Violations inc. ACF	# Rule Violations exl. ACF	Rule Violations A-D	D-C	C-G	ACF	Transitional Probabilities (TPs)
ACFCG	0	0	0	0	0	0	0.52
ADCFC	0	0	0	0	0	0	0.52
ACGFC	*0*	*0*	*0*	*0*	*0*	*0*	*0.43*
ADCGF	*0*	*0*	*0*	*0*	*0*	*0*	*0.57*
ADCFG	1	1	1	0	1	0	0.49
ADFCG	2	1	1	1	0	1	0.45
ADGCF	2	2	2	1	1	0	0.25
ADGFC	3	2	2	1	1	1	0.33
AGCFG	2	2	2	0	2	0	0.19
AGFGC	3	2	2	0	2	1	0.10
AGDCF	3	3	1	1	1	0	0.35
AGFDC	4	3	1	1	1	1	0.30

Fig 6.3. Artificial grammar and sequences. (A) Transition map of the artificial gramma (AG) used to generate sequences in the experiments. (B) 8 exposure sequences generated from the AG. (C) Sequences which were either 'consistent' with the AG in blue or 'violated' the AG (below the black dotted line). Image sourced from Alice Milne's PhD thesis.

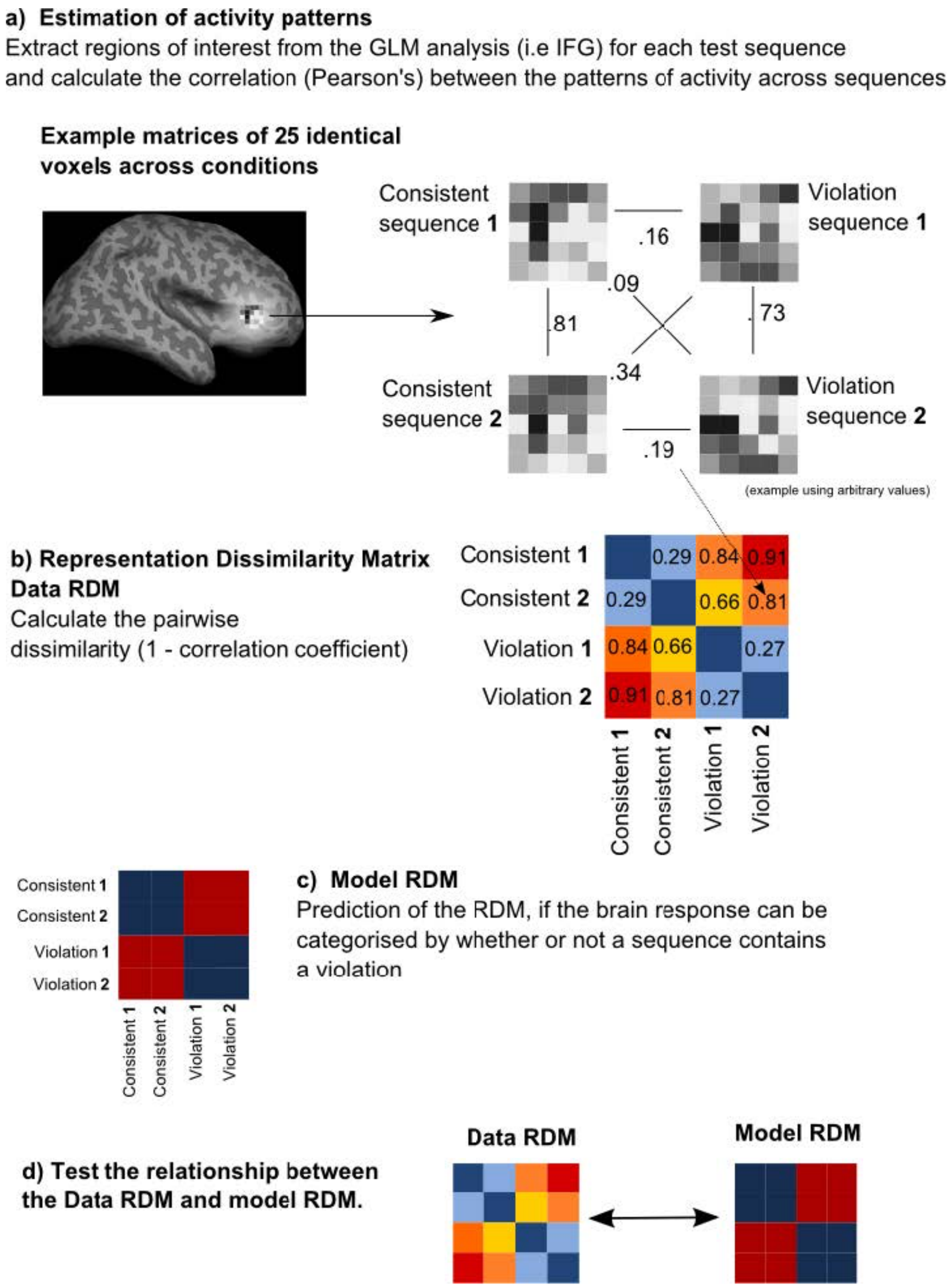
6.5.4 MRI Procedure

The MRI procedure was conducted in the dedicated primate area with a 4.7T Bruker scanner (Bruker, BioSpec 47/60 VAS Bruker Medical, Etlingen, Germany). The Bruker scanner includes a 4-channel surface coil array and a saddle transmitter coil (WK+S, San Diego, California). During the scan, the monkey was positioned in a customise-designed primate chair to fit into the bore of the MRI scanner. An initial fast low angle shot (FLASH) sequence was used to position the functional and anatomical sequences. The functional images were acquired with a gradient-echo echo planar sequence (GE-EPI; TR = 10000ms, TA = 2600ms, TE = 21ms, flip angle = 90º, voxel matrix = 92x92, field of view = 11cm x 11cm x 3.84cm, slice thickness = 1.2mm with no gap, in plane resolution = 1.2mm, number of slices = 48). Anatomical images were acquired using a T1-weighted isotropic magnetization prepared rapid gradient echo (MP-RAGE; TE = 3.8ms; TR=19ms, TA = 2000ms; 96 slices; field of view = 10.7cm x 10.7cm x 5.93cm, voxel matrix = 176 x 176, in-plane resolution = 0.6mm, slice thickness 0.62mm).

The MRI procedure was the same in both the auditory and visual modalities. During the exposure phase, the monkey was presented with 8 repetitions at least with no specific orders of the sequences to get familiar with the sequences that followed the AV Grammar, and scanning was not proceeded during this period. In the meantime, the juice reward was randomly given to the monkey to maintain the motivation. During the testing phase, the functional data was collected with the presentation of testing sequences. In each of the functional runs, there were 38 trials in total with 32 stimulus trials which contained 16 consistent and 16 violation trials, also, 6 silent trials were included. Sparse imaging was used to reduce the impact of the sound noise generated by the scanner during the scan (Kayser, Petkov, Remedios, & Logothetis, 2012). The acquisition of the scanning volume started from the middle point of the test sequence for accounting for the lag of the haemodynamic response (Baumann et al., 2010; Boynton, Engel, Glover, & Heeger, 1996). The juice reward was given to the monkey regardless of how the monkey performed on the eye tracking task in each of the trials. Each scanning session contained 6 runs in general as formerly highlighted, which also depends on the monkey’s behaviour and/or whether the scanner worked properly. Considering of the animal welfare, there might be some cases that the monkey was not able to stay stable in the scanning chair which might result in stopping the scanning session and bringing the monkey back to the colony.

6.6 fMRI RSA Analysis Overview

Representational similarity analysis (RSA) can be utilised to compare the similarity of the multi voxel brain response across audio and visual modalities (Nili et al., 2014). It can also show the behavioural motivation models that fit the data. In each condition (Figure 6.4a), the standardised beta coefficients for all voxels in a region are extracted, and the correlations in pair are calculated for each pair of the conditions. To show the dissimilarity (1-correlation) in pair in all conditions (consistent or violation), a representational dissimilarity matrix (RDM) is generated (Figure 6.4b), and it is referred to as the data RDM here. The further the consistent or violation sequence is, the higher the dissimilarity it would be. For example, consistent 1 versus violation 2, having a dissimilarity of 0.91. It is possible to predict the brain activity by Model RDMS which are generated on the basis of the features of the test sequences. For instance, the binary model (Figure 6.4c) is supposed to accurately execute the prediction of brain activity if the brain response is only based on the presence or absence of a violation. But behaviour is potentially more interesting as might the brain responses to audio-visual sequences. For identifying if any of these models can explain the variance in the brain, we can conduct the test with RSA between the brain activity (data RDM) and predictive models (Model RDMs) (Figure 6.4d).



	Consistent 1	Consistent 2	Violation 1	Violation 2
Consistent 1		0.29	0.84	0.91
Consistent 2	0.29		0.66	0.81
Violation 1	0.84	0.66		0.27
Violation 2	0.91	0.81	0.27	

Fig 6.4. Illustration of RSA procedure analysis. (a &b) Dissimilarity matrices created on the basis of standardised beta coefficients from General Linear Model (GLM) analysis (data RDM). (c) Dissimilarity matrix on the basis of properties of the sequence (Model RDM). (d) Comparisons between Model and data RDMs. Image sourced from Alice Milne’s PhD thesis.

6.7 Pre-processing and Data Analysis

The fMRI Expert Analysis Tool (FEAT, version: 6) was used for the data analysis including the pre-processing. The General Linear Model analysis was also performed in the first level analysis with fixed effects. Notably, the group analysis that had been run partially was conducted locally with FEAT (version: 6.0.6.4). The explanatory variables (EVs) were the sequences that were consistent or violated in each case. The FEAT analysis also involved head motion correction and registration of the functional images to the monkey's own high resolution anatomical scan. Non-brain related tissues or materials were removed with the BET brain extraction. The functional images were registered to a standard space image and the functional EPI images were registered to the high-resolution anatomical image. Finally, the image transformation matrices were concatenated and applied to the functional data to ensure the first-level images being registered to the standard macaque brain.

Regarding the high-level analysis, the individual sessions were grouped using either fixed-effects or mixed-effects with automatic outlier de-weighting. The default setting of the Z-statistics is cluster corrected using a threshold of $Z > 2.3$. The analysis I conducted was based on an exploratory uncorrected threshold with $p < 0.001$. The EVs were set to Grammatical and Ungrammatical in the full model setup. By setting the contrasts with different conditions and modalities, the multiple comparisons were presented with responsive voxels in the brain.

6.8 Results

All the pre-analysis of the primate fMRI data and some visual data has been completed, and below is an example audio/visual fMRI data that has been pre-analysed by me to see which parts of the macaque brain process the structure of auditory/visual sequences.

The figure (Figure 6.5.) below shows three single sessions that are uncorrected for multiple comparisons, and these sessions are presented in coronal, sagittal and axial planes. Panel A shows the visual responsive voxels in inferior temporal visual areas in the visual condition (Picture vs No Picture). The brain activity in auditory cortex was also found as is shown in Panel B in the condition of Sound versus Silence. On top of that, the responsive voxels are found in prefrontal cortex and infra parietal lobule in the contrast of Ungrammatical versus Grammatical, and the sequences were presented violated or followed the rules of the Artificial Grammar. These brain areas show sensitivity to auditory/visual sequences with different contrasts. Additionally, the expected voxels in prefrontal areas on the Grammaticality contrast are similar to what Wilson et al. reported (Wilson et al., 2015) which is encouraging. Group analysis was partially completed as we got some technical issues during the process. The next step in the future would be registering the datasets and conducting a univariate group analysis of the datasets in each monkey and those in the human participants finishing with the RSA analysis as highlighted above.

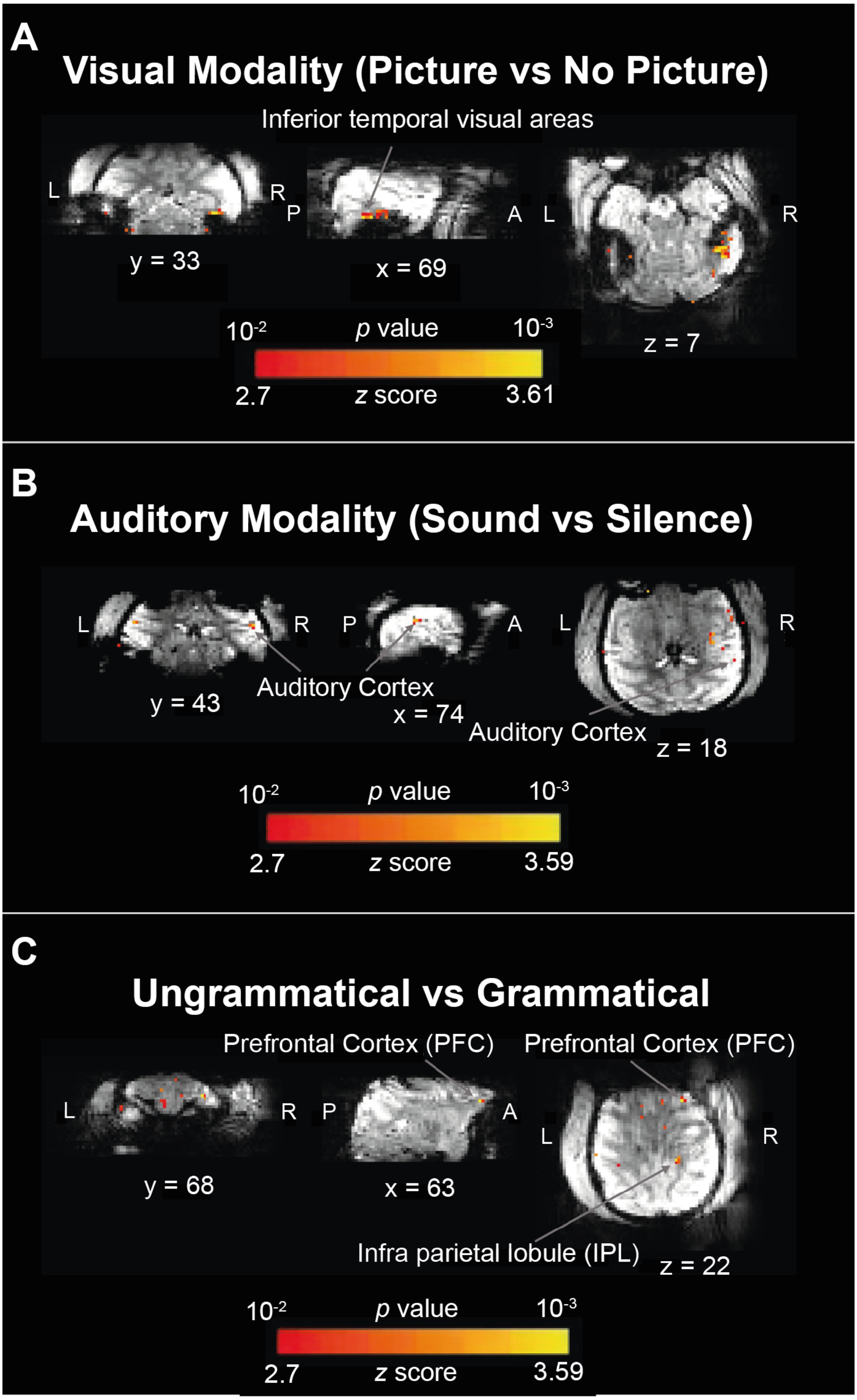


Figure 6.5. Whole brain activity of Monkey 2 across modalities being presented in coronal, sagittal and axial planes. (A) Visual Modality: Picture vs No Picture. An example brain activity

of visual modality under the condition picture vs no picture. It shows activities in the inferior temporal visual areas as pointed in panel (A) in the visual condition. (B) Auditory Modality: Sound vs Silence. An example session of the auditory modality showing the brain activity in auditory cortex in the sound condition. (C) Ungrammatical vs Grammatical. Brain activities in prefrontal cortex and infra parietal lobule in grammatical condition. A (anterior), P (posterior), L (left), R (right).

6.9 Discussion

This comparative fMRI study in humans and macaques aims to investigate how the primate brain processes auditory and visual sequences that have the same structure over time. Specifically, to investigate whether brain areas such as prefrontal cortex and STS will show similar patterns regardless of whether they are auditory or visual sequences, or whether the sensory system is modality-specific when it comes to multisensory integration. The preliminary result shows corresponding auditory and visual responses in different brain areas, and the responses in prefrontal areas show similarity to previous findings under the condition of Ungrammatical versus Grammatical (Wilson et al., 2015). These results also provide evidence that regions of auditory/visual cortex and prefrontal cortex are functionally engaged with auditory/visual sequences that violate or follow the Artificial Grammar.

Although the preliminary result is not showing very robust responses in the brain, it provides evidence of brain areas engaged in auditory and visual sequence processing by using the AG. It is also promising by showing similar findings in the Grammaticality contrast in prefrontal cortex (Wilson et al., 2015). There might be several reasons leading to the weak responsive voxels. Firstly, these are single sessions that are uncorrected for multiple comparisons, so further analysis needs to be processed to investigate more about how the brain processes the auditory and visual sequences that have the same structure. Also, I have tried to set different parameters during the analysis, however, it might be possible that the parameters used were not the most optimised, which said that it might be worth trying to conduct further exploration for future analysis by adjusting the parameters. Additionally, the inherent differences of the monkeys might also be a factor, leading to the varied activations that can be observed or presented. It is unavoidable that the brain is shaped differently in each of the individuals such as the brain size, or the deep structures of the brain such as hippocampus or thalamus. Therefore, it might be worth trying to use different ways for the registration by registering the functional images to the monkey's own anatomical image rather than to a standard template of the macaque brain. By doing this, it might allow us to see some strong activations of the brain areas as the functional scans should align and overlay with the anatomical image better since the differences are minimised because of using the monkey's own anatomical image.

It is also unavoidable that the monkey might shake a lot or not being able to maintain the stability in some of the scanning sessions. In this case, the data collected in these scanning sessions might not be as good as the data collected in other ones (e.g., the noise occurred when the monkey shakes or not looking into the fixation spot on the screen and ignoring the sequences being presented), that the monkey participated well. We could try several ways to minimise the impact by postponing the scans if the monkey still could not stay stable within a few continuous scanning sessions, and this would also be helpful when considering of the animal welfare. Alternatively, we could take the monkey to the laboratory to re-train on the visual fixation task as previously mentioned if necessary. By doing this, it might help to reduce the uncertainties which might affect the data collection during the scans.

Although the preliminary result here is not informative or evident enough to validate the hypothesis proposed above, it is encouraging that the result shows auditory or visual responsive voxels accordingly in the audio/visual condition. Further analysis would be needed to complete the fMRI study, and future studies by using electrophysiological techniques (e.g., electroencephalography) altogether with fMRI and behavioural recordings may provide novel insights and tease apart contributions of the processes involved in multisensory integration (Attaheri et al., 2015; Scanziani & Häusser, 2009).

To summarise, the preliminary result suggests that humans and nonhuman primates appear to share multisensory integration abilities. Further data analysis will be conducted along with the results being generated, which will provide strong evidence and validate the hypothesis proposed above. These are being further progressed including what has been formerly mentioned to complete the comparative human and nonhuman primate neurobiological study.

Chapter 7: General Discussion

Humans can flexibly create complex mental constructions of meaningful information from any sensory input on the dependence of language. It is distinguished from humans that NHPs do not communicate with the full combinatorial capacity of language, however, their cognitive system is known to be sophisticated while processing the integration of various forms of information from any sensory input into meaningful representation. In this thesis, I aimed to determine whether macaque monkeys are capable of combining information from a sequence of two sounds to identify visual objects by colour and shape properties in order to better understand language evolution in relation to combinatorial abilities.

7.1 The Feasibility of the Touchscreen Paradigm

In Chapter 2, I demonstrated the specifics of the touchscreen training methods and task, designing the touchscreen device and modifying the experimental scripts in combination with the consideration of the animal welfare to make sure a smooth testing among the monkeys. There has been a study conducted in humans with similar paradigm finding that humans are able to show generalisation to novel sentences and scenes which they had no experience before in relation to semantics and syntax interface (Poletiek et al., 2021). When it comes to the touchscreen task conducted in NHPs, the first thought might be whether they are able to perform the task. In our study, it has been proven that the touchscreen paradigm worked well when being applied to macaque monkeys, and they were able to learn the task by themselves.

In Chapter 3, I listed the abundant behavioural data collected from two monkeys in different phases and evaluated their performance using various methods. Additionally, I set the criteria to determine whether the monkeys could proceed the following phases of testing. Apart from that, I made huge efforts to improve the touchscreen paradigm to improve the monkeys' performance including introducing distractors to the monkey, optimising the reward and timeouts in the experimental code, examining the response positions to correct monkeys' preferences to any positions within the quadrant, and testing the monkeys in different time periods such as mornings and afternoons to find out the best time for them to get involved in the touchscreen task. Although not all of the improvements worked, I tried to find the most effective way to make sure that the touchscreen paradigm could work the best in monkeys, to give them the best chance to learn this complex task.

7.2 Improvements of Touchscreen Paradigm

7.2.1 Advancements of the Touchscreen Device

On top of that, I tested the monkeys with 3 learning phases with gradually increased complexity, allowing them to have sufficient time to learn the touchscreen paradigm step by step while not testing them without enough gradual process (Medam & Fagot, 2016). There were many difficulties occurring during different learning phases such as the limited access to the touchscreen. More specifically, the monkeys seemed to have the preference of using the right hand to touch the screen. Also, they preferred to touch the right side of the screen even the screen has been divided into 4 equal quadrants. It might be easier for them to touch the right side of the screen which might have some impact on their behavioural strategy. Although the visual stimuli are randomly presented on the screen in different locations, and the stimulus can be presented in different positions even within the same quadrant. If the monkey has a preference of the positions that it would like to select among the 4 quadrants, the reposes could be affected no matter if its correct or not. For example, if the stimulus is presented on the left bottom of the screen, the monkey might ignore this trial and touch anywhere of the screen to complete the trial. In this case, the trial would be an incorrect trial as the monkey does not pick the correct stimulus associated with the sound that has been played, which later might have some effects to the monkeys' performance in average. Therefore, I have also tried to make advancements of the touchscreen device as mentioned before. The straight bars of the home-cage have limited the monkeys' access to the touchscreen when performing the task, we cut the edges of the touchscreen device to make it fit the bottom crush of the home-cage. Additionally, we have made some poles to support the stability of the device. By doing this, there would be no limit when the monkeys try to touch the screen. Although I did not have a chance to test the monkeys on the updated touchscreen apparatus due to many unexpected issues, it could still be used in similar studies in the future. There is no doubt that the touchscreen device could still be improved in relation to various types of rewards to give monkeys more options and associated reward for consistent good performance such as the automatic feeder designed for animal experiments (Oh, Hofer, & Fitch, 2017). Given the fact that time and technical skills were limited, I did not have a chance to completely work on the build of this system to improve it. However, it is still worth trying to make the new touchscreen system in the future.

7.2.2 Efforts to Avoid Bias towards Visual Stimuli

There was a study conducted in NHPs finding that monkeys are able to discriminate one monkey from the others when presenting the stimuli with drawing of lines (Dittrich, 1994). Not only was the task easy enough to be performed by the monkeys, but also that the monkeys are able to perform similar to cognitive tasks the laboratory had tested previously in macaques such as the audio-visual selective attention task (Rinne et al., 2017). Additionally, it was also found in monkeys that they are capable of performing a range of the audio-visual recognition task (Evans et al., 2005). These studies suggest that NHPs can perform even more complex tasks that we still do not understand the limits of their cognitive abilities. It is also possible that it takes substantial amounts of time for the monkeys to understand the task and to work their way through the different phases of learning. This is also one of the reasons of why we tested the monkeys with the touchscreen task for such a long time to ensure that they had as much opportunity to learn as possible. Moreover, many studies have shown that monkeys have either bias to shapes or colours, or tunes, even any of the dimensions involved in their experiments (D'Amato & Salmon, 1982; Fehring et al., 2022; Found & Müller, 1996; Ghasemian et al., 2021b; Kerzel & Schönhammer, 2013; Vázquez et al., 2000; Vonk & Rastogi, 2019). It is also interesting that children seemed to have a shape bias while apes seemed to show the bias to colours (Vonk & Rastogi, 2019). In our study, the monkeys seemed to have slight bias toward the shape at the very beginning of the colour and shape alternation, however, in some of the sessions, the monkeys could also show bias towards the colour. This might be because of the unexpectable behavioural strategy of the monkeys leading to the contradictory findings. Nevertheless, we did not expect the monkeys to show any of the bias to the visual stimuli as in the combinatorial learning phase, relying on either the shape or colour might not be the good indication of combinatorial learning. This is also the reason of spending much time on alternating the colour and shape versions before we step into the final testing phase. We tried to manage the balance between the shape and colour properties in step-by-step training periods. The results in Chapter 4 indicate that the monkey did not show any bias towards the shape or colour. Furthermore, when investigated whether the monkeys had any preference to the visual stimuli, we found that there were slight differences between the stimuli that the monkeys selected in both shape and colour properties. By doing this, any possibility of the preference of visual stimuli was minimised to avoid any impact to the monkeys' performance.

7.2.3 Reducing the Number of Stimuli and Introducing the Foil Distractors

Regarding the number of stimuli, M1 was tested with the more cognitively effortful 3 auditory stimuli in association with 3 visual stimuli at the beginning of the touchscreen learning. We found slight differences between the frequency of the shapes that the monkey selected, which was one of the shapes was not selected as frequently as the other twos (rectangle). Apart from that, the colour purple was not frequently selected as many times as the other two when it came to the colour property. It might be possible that the monkey might pay more attention to unregular objects as the other two were more unregular in the shape dimension when compared to rectangle. Rather than darker colours presented on the black screen, the monkey preferred brighter ones as the colour purple was not selected much often. Considering of improving the monkey's performance as we did not find improvements over time at the initial learning phase, we firstly tried to use 3 extra shapes with no associations to the auditory stimuli, meaning that the extra shapes were distractors. There was no sound being played when any of the 3 shapes was presented on the screen and no juice reward when the monkey selected the distractors. By doing this, we aimed to see whether the monkey was able to discriminate the distractors to the visual stimuli, and we found that the monkey quickly learned to ignore the distractors.

Further methods that I tried to increase the monkey's performance during the initial training phase were to introduce 3 extra auditory stimuli that were distinguished from the ones associated with the visual stimuli. Additionally, the juice reward would not be given no matter whether the monkey made a choice with the correct correspondence or not to the foils. By doing this, we aimed to investigate whether the monkey could learn to ignore the auditory and visual distractors. In other words, theses extra added audio-visual distractors were meaningless just to distract the monkey when tested on the task. If the monkey could ignore the distractors and just select among the stimuli, it could be an indication of that the monkey was able to learn and understand the associations between the auditory and visual stimuli. We found that the monkey still quickly learned to ignore the audio-visual distractors without touching them when the distractors were used for a short period, but the monkey's performance was not increased. Apparently, introducing extra distractors did not help much with the increase of the monkey's performance, we decided not to use them only if needed. We did not see the significant increase of the monkey's performance across sessions, thus, to see whether there was an increasing learning over time of M1 when tested with 3 stimuli as was shown in Chapter4, it was found that the monkey performed better on the shape trials over time even though it was not significantly increased. However, the learning of colour trials and the learning of trials in average showed an opposite way. This contradictory finding suggests that it was difficult for the monkey to maintain a stable performance within a testing session.

Moreover, the number of the visual stimuli was reduced from 3 to 2 for trying to increase the monkey's performance. Although this did not help much either in terms of improving the mean performance, it was one of the many methods we have tried. Also, the monkey did not select one of the shapes or colours as many times as the other ones, this was also one of the reasons that we removed one of the visual stimuli in the shape and colour properties. However, the number of breach differences and the performance for breach trials seemed to be increased as have been shown in Chapter 4 when alternating the colour and shape versions within the same day in both monkeys. This might suggest that reducing the number of stimuli has helped with increasing the monkey's performance. Although there were not many combinations of the shapes and colours when it came to the final testing phase compared to 3 different shapes and colours, the complexity of the task was already difficult to monkeys. More specifically, the monkeys not only needed to learn the association between the sounds and the visual objects, but also had to learn the combinations of the sounds as well as the integrated colours and shapes.

7.2.4 Experimental Environment

In relation to the experimental environment, it might be possible that the monkey could be affected by many factors such as people walking by, and the other monkeys being taken out from their home-cage or being fed. This might also suggest that the experimental condition might make some impacts on the monkey's performance. For example, the baboons involved in the paper of Medam et al. had 24/7 access to the touchscreen device, and they were freely housed in an open area (Medam & Fagot, 2016). This has extensively increased the time of the access to the touchscreen. Considering of the animal welfare and the requirements in our colony, the touchscreen could not be left in the colony overnight. Additionally, using characteristically designed 'testing box' might provide the animals with more quiet experimental environment (Calapai et al., 2022). By doing this, it might be possible that the monkey could pay more attention to the touchscreen task and the time of the access to the touchscreen could be increased in an effective way, making it possible that the monkey's performance could be increased.

7.2.5 Individual Differences between Monkeys

Moreover, the individual differences between each monkey might not be ignored as when we tried to improve the testing parameters in the experimental code such as extending the delay between the trials when the monkey made an incorrect choice, it was found that M1 was not bothered that much compared to M2. Specifically, M2 would just leave the touchscreen there without any involvement when being tested, and M1 could even wait in front of the touchscreen for 3 or more minutes. This indicates that the parameters between individual monkeys could be adjusted to which suits the monkey the best. By doing this, it might help with any of the types of the behavioural studies as individual differences could be an important factor that affect the experiments.

7.2.6 Balance of Shape and Colour Properties

Furthermore, it is important to balance the colour and shape only performance before we step into the final combinatorial leaning phase. As if the monkey has a bias toward the colour or shape property, it might be possible that the monkey just relies on one of the properties while not indeed combing or integrating the colours and shapes. It was found in many studies that NHPs or children showed bias to either the colour or shapes even other dimensions (D’Amato & Salmon, 1982; Fehring et al., 2022; Found & Müller, 1996; Ghasemian et al., 2021b; Kerzel & Schönhammer, 2013; Vázquez et al., 2000; Vonk & Rastogi, 2019). However, the monkeys involved in our touchscreen task did not show significant bias towards either to colours or shapes as was shown in Chapter 4 when switching within the colour and shape only versions. The reason for doing this was to make sure that the monkey was able to integrate the colour and shape rather than relying on just one of the properties. There was slightly bias towards the shapes or colours when we started the touchscreen task, but it was occasionally found within some sessions as the monkeys preferred selecting the colours and performed well in some sessions, while in some other sessions, the monkeys would be more involved when shape stimuli were presented on the screen and showed better performance compared to colours. Our findings in the alternation between colour and shape only versions might suggest that the monkeys’ bias towards either colours or shapes could be corrected when necessary.

7.3 Evidence of Combinatorial Learning

One of the monkeys was progressed to the combinatorial learning phase as another one had to be terminated due to some animal welfare issues. In Chapter 5, M1 showed significant learning in all types of trials including shape only, colour only and combined trials in the familiar condition. In comparison with the two phases tested before, the monkey showed a significant increase of the mean performance (percent correct) across sessions. Additionally, our findings indicate that the monkey was able to learn to identify an object by its joint colour and shape properties. Many other studies have shown that NHPs are able to integrate vocalisations into meaningful ways with minimal combinatorial operation and some monkeys also shown corresponding response to the combinations of alarm calls (Arnold & Zuberbühler, 2006; R. M. Seyfarth et al., 1980). In our study, the monkey was able to listen to a sound that refers to either a shape or colour, and by hearing the combination of two sounds referring to an object with corresponding colour and shape, the monkey could select the corresponding visual object which suggests that common combinatorial semantics ability might be shared with humans. Additionally, the comparisons of the combined trials with the shape foils and colour foils were significant, meaning that the monkey was able to distinguish the colour and shape property, and to distinguish the correct objects from foils that share one of the features. Similar findings were found in the paper that Medam et al. conducted, of which one of baboons was able to learn the compounds of different shapes and colours (Medam & Fagot, 2016). Although they were using the labels of different letters that associated with either the shapes or colours, meaning that no auditory stimulus was used, the monkey could conduct the combinatorial learning entirely in the visual domain. My thesis work obtained evidence for combinatorial learning in the one macaque tested using an audio-visual task.

Although the monkey that I tested did not appear to easily generalise the learned stimuli to novel ones. Namely, when novel combinations of the sounds were presented without reward, the monkey appeared to change its response strategy to these novel sound/object trials, making it difficult to understand whether it would have been able to generalise its learning to novel combinations not previously paired of the stimuli that the monkey had learned.

One possibility is that I knew I had limited time to test the macaque so although we would have preferred to have fewer probe trials (less than 10%) so that the monkey did not change its response strategy by noticing the novel stimulus conditions, we opted for 25% probe trials which seemed to have changed the monkey's behavioural response strategy (performing well below chance on many of the novel tested conditions). Another possibility could be that a few more sessions were needed for the monkey to show the generalisation to novel combinations. Moreover, the novel trials were presented at the relatively high proportion (25%) and were used as probe trials without giving the juice reward. This means that the monkey would not receive the juice reward from the novel trials, regardless of whether the monkey made a correct selection, and this potentially resulted in a change in its response strategy during these trials to not respond to those stimuli at all or to treat them differently to the ones that could lead to reward. This might have impacted the behavioural strategy of the monkey as the absence of a reward could lead the monkey to ignore the trial and touch any stimulus without considering if it was the one associated with the sounds that had been played.

Studies in humans such as Poletiek et al found that humans took only few hours to learn with the similar paradigm I tested on monkeys (Poletiek et al., 2021). The monkeys I worked with by comparison took much longer to progress through the stages of the task, although the comparison cannot be direct because the human participants in the Poletiek study were instructed on how to conduct the task via language, giving them a massive advantage on how to do the task. Again by comparison, the monkey training needed longer periods of time to guide the monkeys through operant training with reward on how to conduct the various stages of the task. In the Poletiek study, both a greater number of auditory and visual stimuli were used, and participants were tested with several experiments accompanied by increased complexity including multiple colour and shape sounds referring to multiple objects. It was also found that the participants in their study showed generalisation to novel sentences and scenes. In contrast, the paradigm I tested on the monkeys does not require the monkeys to learn the complex hierarchical structures that humans could learn with ease, but with simplified audio/visual associative learning and the combinatorial learning at a very basic two sound sequence and one object with distinct colour and shape properties. These simplified testing phases still took the monkey(s) a longer time to learn compared to humans, I, however, also tested the monkeys for longer periods than required for immediate progression when they achieved criterion. For instance, I wanted to make sure that their associative learning was as strong as possible before continuing to the next phase of testing. We are not clear whether it is easier for the monkeys to learn combinatorial semantics if they are exposed to similar environments as humans are, or whether it will help to decrease the time of learning with similar behavioural paradigms. However, the current study provides insights into combinatorial semantics of language learning which might be evolutionarily conserved, with the caveat that the monkeys took longer to learn the basic stages of colour and shape learning. New paradigms have developed for training monkeys on auditory tasks that could take much less time to train (Wikman, Rinne, & Petkov, 2019), but would need to be adapted to the paradigm tested here.

7.4 Future Work and Plan

In the future, the task could be taken further than I have here by testing more monkeys. However, my work has laid the groundwork for more rapid progression of future monkeys working on the task. It also demonstrates the first evidence that a macaque can learn the auditory-visual combinations with reinforcement learning. The monkeys can be possibly tested with advanced apparatus in the future as I mentioned before, without liming their access to the touchscreen. Alternatively, the monkeys can be tested in a characteristically designed box which might ensure a quiet experimental environment, allowing them to focus on the touchscreen task. It might also worth trying to give them different kinds of rewards they prefer according to the number of consecutive correct trials they make to encourage them to do more correct trials. Additionally, it might also be useful to try to design more options of the auditory and visual stimuli which can possibly improve the monkeys' performance. By applying these potential improvements to the touchscreen testing in the future, it is highly possible that the monkeys can perform much better than current thought, and it may save much time on the touchscreen training.

Moreover, I have only conducted the behavioural work, but this behavioural work sets the basis for future neuroimaging studies. Specifically, the touchscreen task could be run in both humans and NHPs during the fMRI scanning. The hypothetic ventral stream of visual processing might start from the V1, going through V2 and V4, and to the IT (Petkov & Jarvis, 2012; Salzmann et al., 2012; Ungerleider, 1982). In parallel, the hypothetic ventral stream of auditory processing in NHPs might start from A1, going through A2 and A3 (Petkov & Jarvis, 2012; Pickles, 1998), the vlPFC might also be involved in the auditory processing to visual integration and processing (Perrodin et al., 2015). In the meantime, the audio-visual sensory convergence might occur in the temporal lobe and the STS in the middle part of the temporal lobe might show robust activities as it is thought to contain conceptual representations of semantic knowledge at least in humans and involved in sensory convergence in both humans and NHPs (Kropotov, 2010;

Perrodin et al., 2015; Specht & Wigglesworth, 2018). On top of that, macaques are a well-established primate model that will continue to make central contributions to neuroscience, but other animal models, such as marmosets, may play complementary scientific roles due to their rich social family network and the way that many animals help in raising the young animals. There has been an increasing attention of using common marmosets in the field of neuroscience based on its small size for more convenient handling and advantages of unique characteristics of behaviour and cognition (Okano, Hikishima, Iriki, & Sasaki, 2012). Apart from that, their fast development, rapid reproduction, and a lifespan that is shorter than macaques but longer than that of mice making them an important animal model for studies as the time of scientific experiments can be shorten (Preuss, 2019). The relative smaller brain compared to macaques and humans of marmosets allows fewer organisational features than other primates, although it is different that the surface of the brain of marmosets is almost free of sulci, their smooth surface of the brain provides advantages on laminar electrode penetration and optical imaging (Mitchell & Leopold, 2015). In some cases, studies of common marmosets might be more accurate compared to rodents due to their high cognitive behaviours underlying neural circuits linking frontal lobe and related subcortical brain areas and mental diseases (Okano et al., 2016). Regarding to the colour sensitive areas in marmosets, colour domains in V1 have tight colocalization with cytochrome oxidase blobs and are closely involved in V2 with thin stripes (Salzmann et al., 2012). Additionally, studies focused on the cone types which are short (S) and long (L) with wavelength-sensitivity also suggest that colour-selective areas in marmosets are V1 and V2 (Mitchell & Leopold, 2015), which is similar to macaques (Salzmann et al., 2012). Additionally, the expanded temporal lobe compared to rodents as well as hierarchically structured sensory cortices means they share a number of similarities with humans, making them a crucial animal model for identifying neural mechanisms that underlies human behaviour and cognition (Fukushima, Ichinohe, & Okano, 2019). Further supported by the fact marmosets share some cognitive similarities with humans due to shared homology (Burkart & Finkenwirth, 2015). Therefore, the proposed fMRI study could possibly be conducted in both humans and NHPs in the future to compare the activities in auditory cortex, visual cortex, and the regions involved in sensory convergence for the assessment of activities in the brain.

Both macaques and marmosets are important for neuroscientific studies as each species contributes unique insights as well as those about evolutionary conservation. Macaques can be used on studies where an evolutionary closer model of human brain function is needed, while marmosets can be used to investigate large group social behaviours and communication in the laboratory. These species of laboratory primates will continue to provide complementary perspectives to answer different aspects of the same overarching scientific question: which aspects of human cognition and language are evolutionary conserved and what is the nature of this system for cognition and communication in nonhuman primates. In our study, one macaque could learn the associations between separate sound and colour/shape and was able to integrate the colour and shape into an object by hearing a short sequence of the sounds referring to the colours and shapes. This ability to use and comprehend the symbols to refer to temporary and spatial objects distantly is thought to be a core feature of human language (Hockett & Hockett, 1960), which suggests that the ability to learn combinatorial semantics might be evolutionarily conserved, pushing back its evolutionary origins to at least a common primate ancestor to humans and Old World Monkeys like macaques that lived approximately 25 million years ago.

The 'Merge operation' is considered to be human unique, which allows humans to conduct the merge operation on linguistic items (words) and apply it to generate an infinite number of expressions using language with more complex hierarchical constructions. One monkey in our study showed that the very basic aspects of this 'Merge operation' might be evolutionarily conserved, which was supported by the evidence that the monkey learned to identify an object by its joint colour and shape when hearing the sounds referring to the colours or shapes. Also, the monkey shows the capability of integrating different colours and shapes into objects with associated sounds, which is similar to human language by combining different words into sentences. This might help to better understand the language processing and cognitive abilities across species. Although it seems unlikely that the monkeys perceived the pairs of sounds as a joint complex sound, because of the limited number of sounds and shapes that the monkeys could manage being tested on, we cannot rule out this possibility that the monkeys memorised all the pairs of sound combinations and associated them with the specific object colour and shape. This however is a less parsimonious explanation because it would appear to be a more cognitively demanding process to memorise all of the sound sequences as single sound objects than to treat each sound as its own object that refers to a particular visual object's colour or shape features.

The current behavioural study may form the basis of a neurobiological or neurophysiological study to better understand neurobiological precursors to language in a nonhuman primate. Thus, by applying the current behavioural study to the neurobiological or neurophysiological study, it might help to understand the language processing in the brain and make contributions to translational neuroscience in humans with language disorders (Cooper, 2001; Laws & Bishop, 2003; Rapin, 2006; Rinaldi et al., 2021).

The behavioural study could for instance form the foundation of a neurobiology study using fMRI in humans and NHPs while performing the combinatorial behavioural paradigm simultaneously in the scanner via moving a lever to make responses or making eye movements to select the visual object after hearing the sequence of sounds. This may provide insights into the comparative neural mechanism and brain function across modalities and across species when studied with comparative neuroimaging or intracranial recording studies in both human neurosurgery patients being monitored for epilepsy and the nonhuman primates as the animal model system. Although the data of the fMRI study was collected, the demands of the touch screen training task, which became more time consuming with the macaques' training than we had anticipated, precluded me from completing the fMRI study analysis within the time required to complete the thesis work. Thus, the completion of the fMRI study of sequence processing of audio and visual modalities across brain regions, which I aim to conduct during my post-doctoral work may provide unique insights into language processing and brain function across species as well.

7.5 Conclusion

The aim of this research is to determine whether macaque monkeys are capable of combining sensory information from sounds to identify visual objects by their joint colour and shape properties in order to better understand language evolution in relation to audio-visual combinatorial abilities. The findings from these experiments indicate that a valuable approach within this field is to include both audio and integrated features of visual objects of the touchscreen learning tasks, providing an important method to assess touchscreen learning capabilities across a range of modalities. The findings provided further detail on the processes that occur during the touchscreen learning, altogether providing novel insights into language evolution whereby one of the monkeys demonstrated audio-visual combinatorial learning at a fundamental level that at a basic level related to Noam Chomsky's Merge operation in human language.

References

Aboitiz, F., & Garcıa, R. (1997). The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. *Brain Research Reviews, 25*(3), 381-396.

Akmajian, A., Farmer, A. K., Bickmore, L., Demers, R. A., & Harnish, R. M. (2017). *Linguistics: An introduction to language and communication*: MIT press.

Arnold, K., & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature, 441*(7091), 303-303. doi:10.1038/441303a

Attaheri, A., Kikuchi, Y., Milne, A. E., Wilson, B., Alter, K., & Petkov, C. I. (2015). EEG potentials associated with artificial grammar learning in the primate brain. *Brain and language, 148*, 74-80.

Baars, B. J., & Gage, N. M. (2010). *Cognition, brain, and consciousness: Introduction to cognitive neuroscience*: Academic Press.

Balezeau, F., Wilson, B., Gallardo, G., Dick, F., Hopkins, W., Anwander, A., . . . Petkov, C. I. (2020). Primate auditory prototype in the evolution of the arcuate fasciculus. *Nature neuroscience, 23*(5), 611-614.

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*(3), 1099-1108.

Belin, P., & Zatorre, R. (2003). Adaptation to speaker's voice in right anterior temporal lobe. NeuroReport: For Rapid Communication of Neuroscience Research, 14 (16), 2105–2109. In.

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

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253-3258.

Bermudez, M. A., & Schultz, W. (2010). Responses of amygdala neurons to positive reward-predicting stimuli depend on background reward (contingency) rather than stimulus-reward pairing (contiguity). *Journal of neurophysiology, 103*(3), 1158-1170.

Berthet, M., Surbeck, M., & Townsend, S. (2025). Extensive compositionality in the vocal system of bonobos. *Science, 388*(6742), 104-108.

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

Binder, M. D., Hirokawa, N., & Windhorst, U. (2009). *Encyclopedia of neuroscience* (Vol. 3166): Springer Berlin, Germany.

Bitar, A. W., Mansour, M. M., & Chehab, A. (2016). *Algorithmic optimizations in the HMAX model targeted for efficient object recognition.* Paper presented at the Computer Vision, Imaging and Computer Graphics Theory and Applications: 10th International Joint Conference, VISIGRAPP 2015, Berlin, Germany, March 11–14, 2015, Revised Selected Papers 10.

Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. *Neuroscience & Biobehavioral Reviews, 47*, 717-734.

Blizzard, S., Fierro-Rojas, A., & Fallah, M. (2017). Response inhibition is facilitated by a change to red over green in the stop signal paradigm. *Frontiers in Human Neuroscience, 10*, 655.

Bondar, I., Vasileva, L., Tereshchenko, L., Buynevich, A., & Latanov, A. (2019). Training of rhesus macaques to complex cognitive tasks. *Neuroscience and Behavioral Physiology, 49*, 996-1007.

Bongioanni, A., Folloni, D., Verhagen, L., Sallet, J., Klein-Flügge, M. C., & Rushworth, M. F. (2021). Activation and disruption of a neural mechanism for novel choice in monkeys. *Nature, 591*(7849), 270-274.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience, 16*(13), 4207-4221.

Brosch, M., Schulz, A., & Scheich, H. (1999). Processing of sound sequences in macaque auditory cortex: response enhancement. *Journal of neurophysiology, 82*(3), 1542-1559.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology, 46*(2), 369-384.

Burkart, J. M., & Finkenwirth, C. (2015). Marmosets as model species in neuroscience and evolutionary anthropology. *Neuroscience research, 93*, 8-19.

Butler, J. L., & Kennerley, S. W. (2019). Mymou: A low-cost, wireless touchscreen system for automated training of nonhuman primates. *Behavior research methods, 51*(6), 2559-2572.

Calapai, A., Cabrera-Moreno, J., Moser, T., & Jeschke, M. (2022). Flexible auditory training, psychophysics, and enrichment of common marmosets with an automated, touchscreen-based system. *Nature Communications, 13*(1), 1648.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in cognitive sciences, 11*(12), 535-543.

Catani, M., Jones, D. K., & Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 57*(1), 8-16.

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

Chomsky, N. (2014). *Aspects of the Theory of Syntax*: MIT press.

Chomsky, N. (2020). Fundamental operations of language: Reflections on optimal design. *Cadernos de Linguística, 1*(1), 01-13.

Chomsky, N., & Skinner, B. F. (1959). Verbal behavior. *Language, 35*(1), 26.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in cognitive sciences, 5*(12), 539-546.

Cooper, J. (2001). Diagnostic and statistical manual of mental disorders (4th edn, text revision)(DSM–IV–TR) Washington, DC: American Psychiatric Association 2000. 943 pp.£ 39.99 (hb). ISBN 0 89042 025 4. *The British Journal of Psychiatry, 179*(1), 85-85.

Crystal, D., & Crystal, D. (1981). Diagnosis and Management. *Clinical Linguistics*, 192-205.

D'Amato, M. R., & Salmon, D. P. (1982). Tune discrimination in monkeys (Cebus apella) and in rats. *Animal Learning & Behavior, 10*(2), 126-134.

Dejerine, J., & Dejerine, J. (1895). *Anatomie des centres nerveux: Méthodes générales d'étude. Embryogénie. Histogénèse et histologie. Anatomie du cerveau* (Vol. 1): Rueff.

DeLillo, C. (1996). The serial organisation of behaviour by non-human primates; an evaluation of experimental paradigms. *Behavioural Brain Research, 81*(1-2), 1-17.

DeValois, R. L., & Jacobs, G. H. (1971). Chapter 3 - Vision1 1This review and the research reported by the writers have been supported by research grants GB-15969 and GB-12303 from the National Science Foundation and EY-00014 from the National Eye Institute, U.S. Public Health Service. In A. M. Schrier & F. Stollnitz (Eds.), *Behavior of Nonhuman Primates* (Vol. 3, pp. 107-157): Elsevier.

Dittrich, W. (1994). How monkeys see others: Discrimination and recognition of monkeys' shape. *Behavioural processes, 33*(1-2), 139-154.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual review of neuroscience, 22*(1), 567-631.

Eckardt, W., & Zuberbühler, K. (2004). Cooperation and competition in two forest monkeys. *Behavioral Ecology, 15*(3), 400-411. doi:10.1093/beheco/arh032

Egelkamp, C. L., & Ross, S. R. (2019). A review of zoo-based cognitive research using touchscreen interfaces. *Zoo biology, 38*(2), 220-235.

Evans, T. A., Howell, S., & Westergaard, G. C. (2005). Auditory-visual cross-modal perception of communicative stimuli in tufted capuchin monkeys (Cebus apella). *Journal of Experimental Psychology: Animal Behavior Processes, 31*(4), 399.

Fehring, D. J., Pascoe, A. J., Haque, Z. Z., Samandra, R., Yokoo, S., Abe, H., . . . Mansouri, F. A. (2022). Dimension of visual information interacts with working memory in monkeys and humans. *Scientific Reports, 12*(1), 5335.

Ferrigno, S., Cheyette, S. J., Piantadosi, S. T., & Cantlon, J. F. (2020). Recursive sequence generation in monkeys, children, US adults, and native Amazonians. *Science Advances, 6*(26), eaaz1002.

Fitch, W., & Jarvis, E. D. (2013). Birdsong and other animal models for human speech, song, and vocal learning.

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a “dimension-weighting” account. *Perception & Psychophysics, 58*(1), 88-101.

Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature neuroscience, 12*(9), 1187.

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

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

Friederici, A. D. (2012). Language development and the ontogeny of the dorsal pathway. *Frontiers in evolutionary neuroscience, 4*, 3.

Friederici, A. D. (2017a). Evolution of the neural language network. *Psychonomic bulletin & review, 24*, 41-47.

Friederici, A. D. (2017b). *Language in our brain: The origins of a uniquely human capacity*: MIT Press.

Friederici, A. D. (2018). The neural basis for human syntax: Broca's area and beyond. *Current opinion in behavioral sciences, 21*, 88-92.

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.

Friedman, H., Ator, N., Haigwood, N., Newsome, W., Allan, J. S., Golos, T. G., . . . Bailey, M. R. (2017). The critical role of nonhuman primates in medical research. *Pathogens & immunity, 2*(3), 352.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in cognitive sciences, 19*(3), 117-125.

Fukushima, M., Ichinohe, N., & Okano, H. (2019). Neuroanatomy of the marmoset. In *The common marmoset in captivity and biomedical research* (pp. 43-62): Elsevier.

Gazes, R. P., Lutz, M. C., Meyer, M. J., Hassett, T. C., & Hampton, R. R. (2019). Influences of demographic, seasonal, and social factors on automated touchscreen computer use by rhesus monkeys (Macaca mulatta) in a large naturalistic group. *PloS one, 14*(4), e0215060.

Gazzaniga, M. S. (2009). The cognitive neurosciences IV. In: MIT press Cambridge, MA.

Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. *Journal of cognitive neuroscience, 24*(3), 564-574.

Ghasemian, S., Vardanjani, M. M., Sheibani, V., & Mansouri, F. A. (2021a). Color-hierarchies in executive control of monkeys' behavior. *American Journal of Primatology, 83*(2), e23231.

Ghasemian, S., Vardanjani, M. M., Sheibani, V., & Mansouri, F. A. (2021b). Dimensional bias and adaptive adjustments in inhibitory control of monkeys. *Animal Cognition, 24*, 815-828.

Ghazanfar, A. A., & Eliades, S. J. (2014). The neurobiology of primate vocal communication. *Current opinion in neurobiology, 28*, 128-135.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in cognitive sciences, 10*(6), 278-285.

Ghazanfar, A. A., & Takahashi, D. Y. (2014). The evolution of speech: vision, rhythm, cooperation. *Trends in cognitive sciences, 18*(10), 543-553.

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*(2), 183-206.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *Journal of Neuroscience, 22*(24), 10829-10837.

Hackett, T. A., Preuss, T. M., & Kaas, J. H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *Journal of Comparative Neurology, 441*(3), 197-222.

Hage, S. R. (2018). Dual neural network model of speech and language evolution: new insights on flexibility of vocal production systems and involvement of frontal cortex. *Current Opinion in Behavioral Sciences, 21*, 80-87.

Haile, T. M., Bohon, K. S., Romero, M. C., & Conway, B. R. (2019). Visual stimulus-driven functional organization of macaque prefrontal cortex. *NeuroImage, 188*, 427-444. doi:https://doi.org/10.1016/j.neuroimage.2018.11.060

Heil, P., Scheich, H., Budinger, E., & Konig, R. (2005). *The Auditory Cortex: A Synthesis of Human and Animal Research*: Psychology Press.

Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American, 203*(3), 88-97.

Hopper, L. M. (2017). Cognitive research in zoos. *Current Opinion in Behavioral Sciences, 16*, 100-110.

Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S., & Wang, L. (2018). Production of supra-regular spatial sequences by macaque monkeys. *Current Biology, 28*(12), 1851-1859. e1854.

Jordan, K. E., Brannon, E. M., Logothetis, N. K., & Ghazanfar, A. A. (2005). Monkeys match the number of voices they hear to the number of faces they see. *Current Biology, 15*(11), 1034-1038.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants′ detection of the sound patterns of words in fluent speech. *Cognitive psychology, 29*(1), 1-23.

Kadia, S. C., & Wang, X. (2003). Spectral integration in A1 of awake primates: neurons with single-and multipeaked tuning characteristics. *Journal of neurophysiology, 89*(3), 1603-1622.

Kanwisher, N., McDermott, J., & Chun, M. M. (2002). The fusiform face area: a module in human extrastriate cortex specialized for face perception.

Kaskan, P. M., Costa, V. D., Eaton, H. P., Zemskova, J. A., Mitz, A. R., Leopold, D. A., . . . Murray, E. A. (2016). Learned Value Shapes Responses to Objects in Frontal and Ventral Stream Networks in Macaque Monkeys. *Cerebral Cortex, 27*(5), 2739-2757. doi:10.1093/cercor/bhw113

Kayser, C., Petkov, C. I., Remedios, R., & Logothetis, N. K. (2012). Multisensory influences on auditory processing: perspectives from fMRI and electrophysiology.

Kerzel, D., & Schönhammer, J. (2013). Salient stimuli capture attention and action. *Attention, Perception, & Psychophysics, 75*, 1633-1643.

Kikuchi, Y., Horwitz, B., Mishkin, M., & Rauschecker, J. P. (2014). Processing of harmonics in the lateral belt of macaque auditory cortex. *Frontiers in Neuroscience, 8*, 204.

Kikuchi, Y., Sedley, W., Griffiths, T. D., & Petkov, C. I. (2018). Evolutionarily conserved neural signatures involved in sequencing predictions and their relevance for language. *Current Opinion in Behavioral Sciences, 21*, 145-153.

Kreifelts, B., Ethofer, T., Shiozawa, T., Grodd, W., & Wildgruber, D. (2009). Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice-and face-sensitive regions in the superior temporal sulcus. *Neuropsychologia, 47*(14), 3059-3066.

Kropotov, J. D. (2010). *Quantitative EEG, event-related potentials and neurotherapy*: Academic Press.

Lameira, A. R., Caneco, B., Kershenbaum, A., Santamaría-Bonfil, G., & Call, J. (2025). Generative vocal plasticity in chimpanzees. *iScience*.

Lanfermann, H., Raab, P., Kretschmann, H.-J., & Weinrich, W. (2019). Cranial neuroimaging and clinical neuroanatomy: atlas of MR imaging and computed tomography. *(No Title)*.

Lattenkamp, E. Z., & Vernes, S. C. (2018). Vocal learning: a language-relevant trait in need of a broad cross-species approach. *Current Opinion in Behavioral Sciences, 21*, 209-215.

Laws, G., & Bishop, D. V. (2003). A comparison of language abilities in adolescents with Down syndrome and children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 46*(6), 1324-1339.

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2010). Multimodal communication by captive chimpanzees (Pan troglodytes). *Animal cognition, 13*, 33-40.

Levelt, W. J. (1999). Models of word production. *Trends in cognitive sciences, 3*(6), 223-232.

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience, 14*(10), 1250-1252.

Liu, D., Chang, Y., Dai, G., Guo, Z., Jones, J. A., Li, T., . . . Wu, X. (2023). Right, but not left, posterior superior temporal gyrus is causally involved in vocal feedback control. *NeuroImage, 278*, 120282.

Mansouri, F. A., Buckley, M. J., Fehring, D. J., & Tanaka, K. (2020). The role of primate prefrontal cortex in bias and shift between visual dimensions. *Cerebral Cortex, 30*(1), 85-99.

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

Mason, S., Premereur, E., Pelekanos, V., Emberton, A., Honess, P., & Mitchell, A. S. (2019). Effective chair training methods for neuroscience research involving rhesus macaques (Macaca mulatta). *Journal of Neuroscience Methods, 317*, 82-93.

Medam, T., & Fagot, J. (2016). Behavioral assessment of combinatorial semantics in baboons (Papio papio). *Behavioural processes, 123*, 54-62.

Milne, A. E., Petkov, C. I., & Wilson, B. (2018). Auditory and visual sequence learning in humans and monkeys using an artificial grammar learning paradigm. *Neuroscience, 389*, 104-117.

Mitchell, J. F., & Leopold, D. A. (2015). The marmoset monkey as a model for visual neuroscience. *Neuroscience research, 93*, 20-46.

Mitchell, J. F., Reynolds, J. H., & Miller, C. T. (2014). Active vision in marmosets: a model system for visual neuroscience. *Journal of Neuroscience, 34*(4), 1183-1194.

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences, 109*(39), 15953-15958.

Nakamura, K., Koba, R., Miwa, M., Yamaguchi, C., Suzuki, H., & Takemoto, A. (2018). A method to train marmosets in visual working memory task and their performance. *Frontiers in behavioral neuroscience, 12*, 46.

Neubert, F.-X., Mars, R. B., Thomas, A. G., Sallet, J., & Rushworth, M. F. (2014). Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron, 81*(3), 700-713.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS computational biology, 10*(4), e1003553.

O'Doherty, J., Rolls, E. T., Francis, S., Bowtell, R., & McGlone, F. (2001). Representation of pleasant and aversive taste in the human brain. *Journal of neurophysiology, 85*(3), 1315-1321.

Oh, J., Hofer, R., & Fitch, W. T. (2017). An open source automatic feeder for animal experiments. *HardwareX, 1*, 13-21.

Okano, H., Hikishima, K., Iriki, A., & Sasaki, E. (2012). *The common marmoset as a novel animal model system for biomedical and neuroscience research applications.* Paper presented at the Seminars in fetal and neonatal medicine.

Okano, H., Sasaki, E., Yamamori, T., Iriki, A., Shimogori, T., Yamaguchi, Y., . . . Miyawaki, A. (2016). Brain/MINDS: A Japanese National Brain Project for Marmoset Neuroscience. *Neuron, 92*(3), 582-590. doi:https://doi.org/10.1016/j.neuron.2016.10.018

Patel, A., Biso, G. M. N. R., & Fowler, J. B. (2022). Neuroanatomy, temporal lobe. In *StatPearls [Internet]*: StatPearls Publishing.

Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature, 439*(7078), 865-870. doi:10.1038/nature04490

Penfield, W., & Roberts, L. (2014). *Speech and brain mechanisms*: Princeton University Press.

Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., . . . Friederici, A. D. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences, 108*(38), 16056-16061.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental brain research, 47*, 329-342.

Perrodin, C., Kayser, C., Abel, T. J., Logothetis, N. K., & Petkov, C. I. (2015). Who is that? Brain networks and mechanisms for identifying individuals. *Trends in cognitive sciences, 19*(12), 783-796.

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

Petkov, C. I., & Wilson, B. (2012). On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1598), 2077-2088.

Petrides, M., & Pandya, D. (1999). Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *European journal of neuroscience, 11*(3), 1011-1036.

Petrides, M., & Pandya, D. (2002). Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *European journal of neuroscience, 16*(2), 291-310.

Phillips, K. A., Bales, K. L., Capitanio, J. P., Conley, A., Czoty, P. W., ‘t Hart, B. A., . . . Nader, M. A. (2014). Why primate models matter. *American journal of primatology, 76*(9), 801-827.

Pickles, J. (1998). An introduction to the physiology of hearing. In *An Introduction to the Physiology of Hearing*: Brill.

Pickles, J. (2013). An introduction to the physiology of hearing. In *An Introduction to the Physiology of Hearing*: Brill.

Poletiek, F. H., Monaghan, P., van de Velde, M., & Bocanegra, B. R. (2021). The semantics-syntax interface: Learning grammatical categories and hierarchical syntactic structure through semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(7), 1141.

Preuss, T. M. (2019). Critique of pure marmoset. *Brain Behavior and Evolution, 93*(2-3), 92-107.

Prinz, W., Aschersleben, G., & Koch, I. (2009). *Cognition and action*: Oxford handbook of human action.

Rapin, I. (2006). Language heterogeneity and regression in the autism spectrum disorders—Overlaps with other childhood language regression syndromes. *Clinical Neuroscience Research, 6*(3-4), 209-218.

Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science, 268*(5207), 111-114.

Rey, A., Perruchet, P., & Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: Evidence from baboons (Papio Papio). *Cognition, 123*(1), 180-184.

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. *Nature neuroscience, 11*(4), 426-428.

Rinaldi, S., Caselli, M. C., Cofelice, V., D'Amico, S., De Cagno, A. G., Della Corte, G., . . . Penge, R. (2021). Efficacy of the treatment of developmental language disorder: A systematic review. *Brain sciences, 11*(3), 407.

Rinne, T., Muers, R. S., Salo, E., Slater, H., & Petkov, C. I. (2017). Functional imaging of audio–visual selective attention in monkeys and humans: How do lapses in monkey performance affect cross-species correspondences? *Cerebral Cortex, 27*(6), 3471-3484.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., & Rabbitt, P. (1994). Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. *Dementia and geriatric cognitive disorders, 5*(5), 266-281.

Roberts, A., Robbins, T., Everitt, B., Jones, G., Wilkinson, J., & Page, K. (1990). The effects of excitotoxic lesions of the basal forebrain on the acquisition, retention and serial reversal of visual discriminations in marmosets. *Neuroscience, 34*(2), 311-329.

Rocchi, F., Oya, H., Balezeau, F., Billig, A. J., Kocsis, Z., Jenison, R. L., . . . Kikuchi, Y. (2021). Common fronto-temporal effective connectivity in humans and monkeys. *Neuron, 109*(5), 852-868. e858.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Salzmann, M. F. V., Bartels, A., Logothetis, N. K., & Schüz, A. (2012). Color blobs in cortical areas V1 and V2 of the new world monkey Callithrix jacchus, revealed by non-differential optical imaging. *Journal of Neuroscience, 32*(23), 7881-7894.

Sapir, E. (1921). An introduction to the study of speech. *Language, 1*, 15.

Scanziani, M., & Häusser, M. (2009). Electrophysiology in the age of light. *Nature, 461*(7266), 930-939.

Schacter, D., Gilbert, D., Wegner, D., & Hood, B. M. (2011). *Psychology: European Edition*: Macmillan International Higher Education.

Schweinberger, S. R., & Burton, A. M. (2003). Covert recognition and the neural system for face processing. *Cortex, 39*(1), 9-30.

Seyfarth, R., & Cheney, D. (2018). Pragmatic flexibility in primate vocal production. *Current Opinion in Behavioral Sciences, 21*, 56-61.

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.

Specht, K., & Wigglesworth, P. (2018). The functional and structural asymmetries of the superior temporal sulcus. *Scandinavian Journal of Psychology, 59*(1), 74-82. doi:10.1111/sjop.12410

Stebbins, W. C. (1971). Chapter 4 - Hearing1 1Preparation of this chapter was supported in part by research grants (NS 05077 and NS 05785) from the National Institute of Neurological Diseases and Stroke. In A. M. Schrier & F. Stollnitz (Eds.), *Behavior of Nonhuman Primates* (Vol. 3, pp. 159-192): Elsevier.

Takemoto, A., Izumi, A., Miwa, M., & Nakamura, K. (2011). Development of a compact and general-purpose experimental apparatus with a touch-sensitive screen for use in evaluating cognitive functions in common marmosets. *Journal of Neuroscience Methods, 199*(1), 82-86.

Tanigawa, H., Lu, H. D., & Roe, A. W. (2010). Functional organization for color and orientation in macaque V4. *Nature neuroscience, 13*(12), 1542-1548.

Terrace, H., & McGonigle, B. (1994). Memory and representation of serial order by children, monkeys, and pigeons. *Current Directions in Psychological Science, 3*(6), 180-185.

Trask, R. L., & Stockwell, P. (2007). *Language and linguistics: The key concepts*: Routledge.

Ungerleider, L. G. (1982). Two cortical visual systems. *Analysis of visual behavior, 549*, chapter 18.

Vázquez, P., Cano, M. n., & Acuña, C. (2000). Discrimination of line orientation in humans and monkeys. *Journal of neurophysiology, 83*(5), 2639-2648.

Vonk, J., & Rastogi, G. (2019). The “sh-ape bias” in non-linguistic categorization: Comparisons between children and other apes. *Journal of Cognition and Development, 20*(3), 380-398.

Wang, L., Amalric, M., Fang, W., Jiang, X., Pallier, C., Figueira, S., . . . Dehaene, S. (2019). Representation of spatial sequences using nested rules in human prefrontal cortex. *NeuroImage, 186*, 245-255.

Wikman, P., Rinne, T., & Petkov, C. I. (2019). Reward cues readily direct monkeys’ auditory performance resulting in broad auditory cortex modulation and interaction with sites along cholinergic and dopaminergic pathways. *Scientific reports, 9*(1), 3055.

Wilson, B., Kikuchi, Y., Sun, L., Hunter, D., Dick, F., Smith, K., . . . Petkov, C. I. (2015). Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nature communications, 6*(1), 8901.

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.

Wise, R. J., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems withinWernicke's area'. *Brain, 124*(1), 83-95.

Yamazaki, Y., Echigo, C., Saiki, M., Inada, M., Watanabe, S., & Iriki, A. (2011). Tool-use learning by common marmosets (Callithrix jacchus). *Experimental brain research, 213*, 63-71.

Zilles, K., & Amunts, K. (2018). Cytoarchitectonic and receptorarchitectonic organization in Broca's region and surrounding cortex. *Current opinion in behavioral sciences, 21*, 93-105.

Zuberbühler, K. (2000). Referential labelling in Diana monkeys. *Animal Behaviour, 59*(5), 917-927. doi:https://doi.org/10.1006/anbe.1999.1317

Zuberbühler, K. (2001). Predator-specific alarm calls in Campbell's monkeys, Cercopithecus campbelli. *Behavioral Ecology and Sociobiology, 50*(5), 414-422. doi:10.1007/s002650100383