Developing Fingerprint Examination Expertise using Simultaneous and Sequential Presentations of Interleaved Practice

Jade Cascun

This thesis is submitted in partial fulfilment of the Honours degree of Bachelor of Psychological Science (Honours)

School of Psychology

The University of Adelaide

September 2020

Word count: 9454

Table of Contents

List of Figures

List of Tables

Abstract

Court practicing fingerprint examiners consistently demonstrate superior performance to novices in their ability to differentiate matching from non-matching fingerprints. Much is known about what differentiates a fingerprint expert from a novice, but less is known about whether the development of fingerprint expertise can be enhanced using empirically based learning strategies. Interleaved practice (mixing the study of one category with the study of another) is an effective strategy for learning visual categories that are highly similar in appearance. Research on interleaving suggests that the sequence in which materials are presented, such as one at a time (Sequentially), or multiple at once (Simultaneously), can influence how well materials are learnt. To date, the benefits of Sequential and Simultaneous interleaving have not been compared in fingerprints. This study employed a 3x11 mixed factorial design to observe learning differences between training tasks (Simultaneous or Sequential fingerprint training, or Control filler task) across 11 sessions. Participants (n=19) were randomly assigned to a training task, and a robust measure of fingerprint expertise, “the Expertise Quotient” (xQ), measured learning across sessions. Results of a mixed factorial ANOVA found no significant interaction between training task and session. Trend analysis revealed that the Sequential fingerprint training condition performed better on the xQ across sessions compared to Controls, though the Simultaneous fingerprint training condition performed worse across sessions in comparison. These results provide important implications for the future development of fingerprint examination training tools. Further research is needed to find the most optimal way to develop fingerprint expertise.

Keywords: perceptual learning, forensic science, fingerprint examination, interleaved practice, visual cognition

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

September 2020

Contribution Statement

My supervisor and I collaborated to formulate a research question and decide on a suitable research methodology. We collaborated to design the computer-based fingerprint training tasks, and myself and another honours student collaborated to arrive at a control task that would be suitable for both of our individual experiments. This means that myself and another honours student used a shared control participant pool. Given it was difficult to recruit participants with social distancing requirements at the time of data collection, pooling control participants together was the only feasible way to safely collect enough data for this experiment. I collected data from 13 out of 19 participants, and data collection for the remaining six participants was collected by my supervisor and another honours student. My supervisor programmed the fingerprint training tasks, control task, and measure of fingerprint expertise to form my data collection tool. My supervisor guided me through using R and conducting analyses in RStudio, and I reproduced the analyses using SPSS. I completed the ethics application, preregistered the details of my experiment on the Open Science Framework, and ran pilot test simulations of the experiment for all participants across each session. I also completed the write-up of this thesis. Lastly, incentives for participation were provided by my supervisor.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Rachel Searston. In honours, they say you either get a great supervisor or a great topic, yet somehow, I got both. Rachel, your excitement about forensics is incredibly inspiring, and your patience to explain statistics to someone as confused as me is something to be admired. I am truly grateful to have learnt even a fraction from the wealth of knowledge that you have. Thank you for everything you have helped me achieve this year. Thank you for being the expert where I was the novice.

To the gals in Hughes 253: “Social Support Wednesdays”, a time designated to studying but usually spent chatting. These days made all the difference to my entire honours experience. Claire and Jonica, thank you for the guaranteed laughter, the space to debrief, the celebration of all wins, and the important reminders to take a break. You both inspire me to do more. I cannot wait to see the things you both go on to achieve in this world.

To my participants, thank you for making this project special. Never did I ever imagine conducting a formal academic experiment at the dinner table. Thank you for your determination to push through those 11 days. Thank you for making this project possible. You are all superstars.

To all the fam at home: Studying interstate is already a daunting thing, and the closed borders of a pandemic can sure throw a spanner into the mix – but your spontaneous phone calls, check-in text messages, and air-mailed care packages were the perfect bundle of support any honours student could ever dream of. Thank you for continuing to support me no matter what crazy idea I come up with.

And to Johny, your endless selflessness, unfailing encouragement, and thoughtful reassurance is something I will forever be grateful for. Thank you for always having confidence in me. I can only ever aspire to be half of the person you are.

Chapter 1: Introduction

When forensic evidence, such as a bitemark, handwriting, shoeprints, or a fingerprint is found at a crime scene, it is sent to a forensic examiner for analysis to uncover more information about the nature of the crime and the people involved (Busey & Parada, 2010; Towler, Ballantyne, Searston, Martire & Kemp, 2018). In many cases, including in the case of fingerprints, examiners analyse the evidence by visually comparing it to similar samples taken from known sources. That is, the examiner, not a computer, decides whether the samples are visually similar enough to conclude whether they were left by the same person (a match) or by different people (Searston & Chin, 2019). Accurate expert judgments about forensic evidence are therefore crucial for identifying genuine suspects in criminal investigations. This honours project investigates perceptual training methods for developing accurate expert judgment in the context of fingerprint examination.

When a fingerprint examiner provides expert testimony in court about whether a pair of prints were left by the same finger, their evidence can be used by judges and juries to help decide whether the accused is guilty or not guilty of the crime at hand. These high-stakes decisions are made even more challenging in cases where the fingerprints were left by the same finger but appear to be different as a result of distortion (friction ridge skin can expand, contract, and shift in numerous directions as it makes contact with different surfaces). Likewise, highly similar, but non-matching prints are increasingly common, also making the examiner's job more challenging (Towler et al., 2018). Thus, the level of performance expected of a practicing fingerprint examiner is understandably high. Fortunately, practicing fingerprint examiners have demonstrated an impressive ability (relative to novices) to distinguish between matching and non-matching pairs of fingerprints (Tangen, Thompson & McCarthy, 2011; Thompson, Tangen & McCarthy, 2014; Ulery, Hicklin, Buscaglia & Roberts, 2011).

Unlike other domains of forensic evidence, fingerprint examiners cannot explicitly show how they arrived at a particular decision, as their analysis of the evidence takes place using a series of complex cognitive processes (Busey & Parada, 2010). These thought processes have been likened to the ‘black box’ metaphor: the input (a pair of fingerprints to be analysed) and output (decision about the fingerprints after analysis) are known, but the mechanisms inside the box (cognitive processes used during analysis) are not (Mnookin, 2008; Searston & Chin, 2019). To accommodate for what cannot be easily shown, cognitive scientists have instead provided fingerprint examiners with the opportunity to demonstrate their expertise, by comparing their performance with the performance of those who do not have past experience analysing fingerprints (Stevenage & Pitfield, 2016; Thompson et al., 2014). As a result of this research, much more is known about what it means to hold expertise in the field, and what differentiates these skills from the average person (Tangen, 2013).

Before considering how trained fingerprint examiners perform in comparison to novices, it is important to understand what having the title of an expert entails. Expert performance has been defined as “…consistently superior performance on a specified set of representative tasks for a domain” (Ericsson & Lehmann, 1996, p. 277). Thus, expertise is more than just the completion of training and experience, but it is also being able to consistently demonstrate skill and competency in the domain that expertise is being claimed. As scientists have put it, “where there is no demonstrably superior performance, there is no expert” (Martire & Edmond, 2017, p. 972). Tests of expertise are often measured under strict laboratory conditions that are more controlled in comparison to what the professionals may be familiar with in the field. For example, research on fingerprint expertise often takes advantage of employing ‘forced choice’ responses to questions in their studies (such as dichotomous ‘same finger’ or ‘different finger’ responses), whereas in the field it is possible to instead draw inconclusive results (Tangen et al., 2011). By pressing experts to select from dichotomous responses instead of allowing them to be indecisive, researchers are better able to ensure their tests are sensitive to capturing genuinely superior performance.

Cognitive scientists have begun assessing fingerprint expertise by using methodology that resembles the identification procedures used in practice, with a few important deviations (Tangen et al., 2011; Thompson, et al., 2014). Typically, both experts and novices are asked to make judgements about pairs of prints presented side-by-side, by deciding if the prints were left by the same finger or different fingers. Unlike real world circumstances, however, the source of all prints used in these studies are known, which means that the number of correct responses can be calculated to make comparisons between the levels of novices’ and experts’ ability. This approach has revealed that experts make significantly less errors in their conclusions about fingerprint pairs than novices (Tangen et al., 2011), and that fingerprint examiners perform particularly well at distinguishing between pairs of prints that are highly similar in appearance, but not a match – a task that novices routinely perform poorly in (Thompson & Tangen, 2014; Thompson, Tangen & McCarthy 2013, 2014; Ulery, Hicklin, Buscaglia & Roberts, 2011). In one study, for example, experts correctly identified 99% of highly similar non-matching prints as a non-match, whereas novices correctly declared just 45% of these cases as non-matches (Tangen et al., 2011). Experts have also demonstrated the capacity to maintain higher levels of accuracy when prints are presented inverted; when the print pairs are separated by a time delay (Thompson & Tangen, 2014); and even when they are only able to glance at the prints for 400 milliseconds (Searston & Tangen, 2017a). Furthermore, the capabilities of fingerprint experts go beyond being able to judge whether or not a print was left by the same finger, experts can also accurately determine whether or not a print was left by the same person, not just the same finger (Searston & Tangen, 2017b). The consistently superior performance displayed by practicing fingerprint examiners in these studies, particularly in comparison to novices, suggests they hold genuine expertise in their field of practice.

Prior research has demonstrated the perceptual expertise of fingerprint examiners and some of the important differences between experts and novices. A natural next step is to take what is known about fingerprint expertise and use this information to benefit the field of practice. One plausible approach is to focus on how to improve the training processes used by new fingerprint examination recruits, so as to boost the development of accurate fingerprint examination ability. Testing empirically based learning strategies in the applied fingerprint domain is an appropriate starting point. Ultimately, knowing how best to optimise learning in the fingerprint domain can form the basis for the development of more efficient, empirically based training tools, and allow for continuous improvements in practice. More widely, this type of research can also contribute to a reduction in the overall number of wrongful convictions that take place in the forensic domain. Given the high-stakes nature of the fingerprint examination profession, research dedicated to enhancing the field of practice is invaluable.

The ability to correctly identify whether a pair of fingerprints are a match is not impossible, and novices do have some ability to do this (Stevenage & Pitfield, 2016; Vokey, Tangen & Cole, 2009). What remains unique to experts, however, is their distinct ability to consistently outperform novices when pairs of prints are highly similar in appearance, but not a match. When humans encounter an unknown object for the first time, they use past examples of similar-looking objects to find a category in memory where the new object best fits, a process referred to as induction (Kellman & Garrigan, 2009; Medin & Schaffer, 1978; Michalski, 1983; Nosofsky, 2011). As people increase their exposure to a particular category, they develop the ability to classify and refine those objects more specifically (Kellman & Garrigan, 2009; Tanaka, Curran & Sheinberg, 2005). For example, most people will have stored in memory their own category for dogs, containing all the different breeds of dogs they have previously seen before. People generally know enough about the broad visual features typically seen in dogs, so even when encountering a rare or unusual mixed-breed, it is similar enough to their established concept of a dog that they are still able to correctly classify it. However, where an avid lover of dogs can easily distinguish a Golden Retriever from a Labrador, someone with less exposure to dogs might see these two breeds as the same. With this in mind, it seems logical that in contrast to a novice, a fingerprint expert who has encountered many more instances of fingerprints, would be more accurate at differentiating between even the most similar of print pairs. What becomes of interest in the present research, then, is how can the process of learning the differences between categories be optimised in order to benefit the development of expertise, particularly within the domain of fingerprints?

Within the field of research on induction, or, learning by example, interleaved practice repeatedly yields the greatest benefit for learning categories that are difficult to distinguish between (i.e., highly similar categories, such as llamas and alpacas). Interleaved practice (or ‘interleaving’), is the process of mixing the study of one category with the study of another, rather than studying each category separately (for example, repeatedly alternating the study of llamas and alpacas, rather than first studying llamas, and then studying alpacas – which is a competing study method known as ‘massing’). Typically, studies have compared the effect of interleaving and massing on induction, by presenting participants with examples of visual stimuli (such as pictures of butterfly species, or images of different painting styles) in either an interleaved or massed sequence (Birnbaum, Kornell, Bjork & Bjork, 2013; Kang & Pashler, 2012; Kornell & Bjork, 2008; Wahlheim, Dunlosky & Jacoby, 2011; Yan, Soderstrom, Seneviratna, Bjork & Bjork, 2017). These studies have then tested the effect of each learning style (interleaved and massed) by having participants complete a subsequent classification task. These tasks involved presenting participants with unseen examples of stimuli from the categories they had just studied, and then asking them to indicate which category they thought the unseen images belonged to. Interleaving consistently showed the greatest benefit on induction compared to massing, though this effect was only found when the categories of stimuli being studied were highly similar to each other in appearance (Carvalho & Goldstone, 2014b).

Naturally, interleaving facilitates the comparison of examples between categories, because the to-be-learnt categories are continuously alternating (Carvalho & Goldstone, 2014; Zukiply & Burt, 2013). When the categories being studied are highly similar, these continuous comparisons highlight the differences between those categories (of which there are few), rather than the similarities between the categories (of which there are many; Zukiply & Burt, 2013). The opportunity to discriminate between categories that interleaving allows is thought to be the reason why interleaving, and not massing, enhances induction for highly similar categories (Birnbaum et al., 2013; Carvalho & Goldstone, 2014b; Kang & Pashler, 2012; Zukiply & Burt, 2013).

When using interleaved practice, few studies have sought to find the most optimal way to present stimuli in order to most benefit learning. Instead, most studies have simply employed a single presentation sequence to represent interleaving, without considering alternative presentation orders (Birnbaum et al., 2013; Kornell & Bjork, 2008). However, among the studies that have considered alternative presentations, it has become apparent that perhaps presenting stimuli from different categories at the same time (simultaneously), can produce a greater learning benefit than presenting stimuli from different categories one at a time (sequentially; Carvalho & Goldstone, 2014a; Kang & Pashler, 2012; Wahlheim et al., 2011).

Of particular interest in the literature is a single study that made direct comparisons between simultaneous and sequential presentations of interleaved learning, using a series of abstract blob-figure stimuli (Carvalho & Goldstone, 2014a). In this study, all participants were presented with a single stimulus on a computer screen and asked to classify it into one of three categories of interest. Their decision was followed by immediate corrective feedback. In the sequential group, this process was repeated until all stimuli had been studied and tested. However, in the novel simultaneous group, the stimuli that had just been tested would move to the left of the screen along with its correct category name and remain there for the duration of the subsequent trial. This process was repeated, with the image and category name of the stimulus that immediately preceded always available to the left of the screen, allowing participants to directly compare the previous stimulus with the current. Performance on the classification task for the simultaneous group was significantly greater than performance in the sequential group, suggesting that there is a benefit to presenting stimuli simultaneously, rather than one at a time. Similarly, Wahlheim et al. (2011) found that the effects of interleaving were enhanced when stimuli from different categories were presented in pairs (simultaneously).

A possible explanation for the benefits of presenting stimuli simultaneously lies in the idea that, rather than having to rely on short-term memory to accurately remember the features of the previous stimulus, presenting stimuli side-by-side allows the categories to be directly compared, making the differences between the categories even more salient (Finley, Roediger, Hughes, Wahlheim & Jacoby, 2015; Rieber, 1966). Although past research suggests that, in abstract blob-figure stimuli, simultaneous presentations are more beneficial to learning than sequential presentations, it cannot be assumed that this advantage will also be found in the fingerprint domain.

Minimal studies have made efforts to apply interleaved practice to the study of fingerprints. However, some studies have used contrast practice techniques similar to interleaving, and found this technique to be advantageous for learning in the fingerprint domain (Searston & Tangen, 2017c). Such research has chosen to test whether participants can learn to distinguish between fingerprints by contrasting matching and non-matching prints. Overall, the contrast practice group of this experiment were better able to differentiate between fingerprints when compared to a baseline study group that did not use any particular learning strategy. Interestingly, this finding was significant in print pairs that were highly similar, but not a match, where the contrast practice training group correctly declared 74% of print pairs, compared to 51% in the baseline group. It is important to note, however, that the focus of this experiment was to compare the effectiveness of different learning strategies, rather than the effectiveness of different methods for presenting stimuli. Thus, only one possible presentation sequence was used. Therefore, while this study has provided evidence for the effectiveness of using a contrasting technique in the study of fingerprints, the most optimal way to present stimuli in the fingerprint domain still remains unknown.

1.1 The Current Experiment

To my knowledge, no studies have directly compared simultaneous and sequential interleaving in fingerprints. However, past research has indeed shown that the benefits of interleaving seen in the study of birds (Birnbaum et al., 2013; Wahlheim et al., 2011), butterflies (Birnbaum et al., 2013), and paintings (Kang & Pashler, 2012; Kornell & Bjork, 2008), might also be applicable to fingerprints (Searston & Tangen, 2017c). The present research will expand on these findings by investigating whether a greater learning benefit is produced when interleaved fingerprints are studied one at a time, or when more than one fingerprint is studied at a time. This study aims to answer whether the development of fingerprint expertise can be enhanced by interleaving fingerprints from different categories on the screen at one time during study.

For the purpose of this experiment, novel computer-based training programs were designed. Participants will take part in multiple sessions of either an interleaved fingerprint training task (Simultaneous or Sequential), or a word scramble task (Control). During these sessions, participants' learning will also be assessed by testing their ability to differentiate between fingerprints using a measure of fingerprint expertise. Based on the success of interleaving in other highly similar categories of stimuli (e.g., Birnbaum et al., 2013; Kornell & Bjork, 2008), it is expected that test scores on a measure of fingerprint expertise will improve over the course of the training sessions. We expect that this improvement will only be seen for participants who complete an interleaved fingerprint training task, but not for participants who complete a Control word-scramble task (Hypothesis 1). Given that simultaneous presentations of stimuli have presented greater learning advantages outside the fingerprint domain (See Carvalho & Goldstone, 2014a), it is also expected that the performance improvement observed in the interleaved fingerprint training conditions will be greater in the Simultaneous condition, compared to the Sequential condition (Hypothesis 2).

Chapter 2: Method

At the early stages of data collection, I preregistered my research plan for this experiment on the Open Science Framework (OSF). The details of my experiment are listed under 'Experiment 2', here. This link includes details of my study, such as data analysis plans, predictions, and methodology.

2.1 Research Design

We employed a 3 (training task: Simultaneous or Sequential fingerprint training, and Control filler task) x 11 (session: Session 1-11) mixed factorial design. Training task was manipulated between-subjects, where each participant was randomly allocated to one of the three training tasks (Simultaneous, Sequential, or Control). Session was manipulated within-subjects, meaning each participant completed all 11 Sessions of their respective training task. The dependent variable in this experiment was participants' mean scores on a standardised test of fingerprint expertise (the Fingerprint Expertise Quotient or fingerprint "xQ") described below.

2.2 Participants

A convenience sample of 19 participants [1] (11 females and eight males) aged between 19-61 (*M* = 34.68, *SD* = 15.60) took part in this study. Due to the social distancing practices in place as a result of COVID-19 at the time of recruitment, participants comprised family and friends of myself, my supervisor, and another honours student, who were all able to complete the experiment remotely. Given the difficulty of recruitment during this time, myself and another honours student collaborated to arrive at a control task that would be appropriate for each of our

---

[1] A coding error was discovered during data collection which resulted in data being collected from an additional participant than planned, who was randomly assigned to the Control condition (n = 7).

individual experiments. For this reason, participants in the Control condition (n = 7) were recruited by both myself and another honours student to form a shared control participant pool. The remaining participants (Simultaneous = 6, Sequential = 6) were recruited solely for my study. Participants were verbally invited to participate, and, given the personal nature of these relationships, it was emphasised that participation was voluntary, and all were informed of their right to withdraw from the study at any point without any repercussions. Four participants volunteered to complete the experiment without compensation. The remaining participants were compensated for their time with a $20 Coles/Myer gift voucher for every accumulated hour of participation (~4 hours).

Participants were required to be 18 years or older, fluent in the English language, and have normal or corrected-to-normal vision. It was also a requirement for participants in this study to have no prior experience working with fingerprints. This requirement was to ensure all participants could be referred to as true novices. Participants who responded in less than 500ms for over 30 percent of trials were to be excluded from the experiment. This project was approved by The University of Adelaide Human Research Ethics Subcommittee (approval number 20/29).

2.3 Power Analysis

In order to determine sample size requirements for this experiment, we conducted an a priori power analysis using Jake Westfall's PANGEA (v0.2) application (Westfall, 2020). Previous studies on interleaved learning typically show effect sizes which range between 0.7 and 1.23 (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008). Accounting for the possibility that the effect sizes reported in the literature might be inflated as a result of publication bias, we aimed to conservatively detect an effect size that is at least as large as the smaller of the reported effect sizes ($d = 0.7$). In order to detect an effect of this size, we calculated that a sample size of at least

18 would be required, with six participants in each condition. Given this sample size, and a total of 34 observations on the measure of fingerprint expertise per participant, per session, we estimated that the experiment would have at least 82.3% statistical power to detect an interaction effect between training task and session.

2.4. Materials

2.4.1 Interleaved Fingerprint Training Tasks. The interleaved fingerprint tasks used in this experiment were custom designed for the purpose of this study. The design of the tasks was inspired by the learning tasks used in previous research on interleaved practice (See Birnbaum et al., 2013; Carvalho & Goldstone, 2014a; Kang & Pashler, 2012). Many of these studies have asked participants to observe images of different stimuli (such as species of birds) and presented the name of the category it belongs to (e.g., ‘finch’) below the image for participants to study and memorise. However, for the present experiment, rather than asking participants to observe stimuli, I considered a more active approach to learning by asking participants to classify the fingerprints being learned into specific categories, similar to the procedures used with abstract blob figure stimuli (See Carvalho & Goldstone, 2014a). We also incorporated an immediate corrective feedback element following each response. Immediate feedback has been found to benefit learning as it validates correct responses (Vojdanoska, Cranney & Newell, 2010), and facilitates a problem-solving attitude when responses are incorrect (Carvalho & Goldstone, 2014a).

2.4.2 Fingerprint Stimuli. The fingerprints used in the interleaved training tasks were sourced from the US National Institute of Science and Technology (NIST) Special database 300. Originally, there were 6,631 fingerprints in total, though some of the images contained descriptive labelling information (such as 'left thumb') which may have impacted the effectiveness of the training if they remained in the image. For this reason, my supervisor applied a 500x500 circular, feathered mask to all prints, removing most of the labelling information. The fingerprint images were then manually sorted through to remove any prints that still contained labelling information, resulting in a total of 5105 fingerprint images to sample from for each session of the fingerprint training tasks. The fingerprint images used in the training tasks are independent of those used in the measure of fingerprint expertise. Thus, any improvement in test scores on the measure of fingerprint expertise cannot be attributed to memorising particular images during training sessions.

2.4.3 Control Word Scramble Stimuli. A total of 1371 common seven letter words (ranging from Abalone – Zealous) were used in the Control condition. These words were sourced from an open-source repository on GitHub, available here.

2.5 Measures

2.5.1 Measures of Fingerprint Expertise: The Expertise Quotient (xQ). In order to measure learning, the xQ measure of fingerprint expertise was completed by all participants at the beginning of every session (Searston, Tangen & Thompson, unpublished; manuscript in preparation). The xQ comprises three different tasks that each capture an element of real-world fingerprint examination processes: A Fingerprint Matching task, a Fingerprint Recognition task, and a Print Nomination task.

In the Fingerprint Matching task, participants are presented with 12 cases of fingerprint pairs, and are asked to determine whether they believe each pair of prints are a match, or not a match. Next, the Fingerprint Recognition task presents 12 trials of a single fingerprint and asks participants to view each print for 10 seconds. Once the 10 seconds has passed, participants are to scroll through an additional 10 prints and select the one that they believe is a match with the print just studied. Lastly, the Print Nomination task requires participants to indicate for a total of 10 prints, whether they believe each print was left by a left or right hand, and from which finger nomination (i.e., thumb, index, middle, ring, or little). Thus, the xQ comprises a total of 34 cases, and takes approximately 12 minutes to complete. The fingerprint materials used in each trial are randomly selected (with replacement) for each participant, each time they take the test. Scores on the xQ are given by calculating the percentage of correct responses in each subtask of the measure. Higher total scores on the xQ indicate a greater ability to discriminate between fingerprints.

The xQ was created by inviting 44 Australian fingerprint experts, and 44 matched novices, to complete 10 tasks that were designed based on different aspects of real-world fingerprint examination processes. The original 10 tasks included 603 cases taking over three hours to complete. The 10 tasks were then reduced down to the three tasks that most optimally discriminated between experts and novices. The length of the xQ (34 cases) was determined by computing all-possible combinations of cases needed to classify experts as experts and novices as novices.

2.6 Software

Participants completed the xQ and training tasks on either a 13-inch or 11-inch MacBook Air laptop. LiveCode Community (version 9.5.1) was used to program the training tasks, run pilot simulations, present the trials to participants, and record their responses. Adobe Photoshop was used to apply the circular feathered masks to the original fingerprint images using an automated batch process. The Apple 'Numbers' application (v10.0) was used to visually inspect pilot simulation data to ensure all information was captured correctly before commencing data collection with participants. All data analyses were conducted using RStudio and reproduced in SPSS.

2.7 Procedure

Prior to commencement of this experiment, participants were provided with an information sheet (Appendix A) and were asked to sign a hardcopy consent form (Appendix B). For each of the 11 sessions, participants were seated individually in a quiet room with a laptop, which was pre-loaded to the experiment home page on LiveCode. On commencement of the first session of this experiment, participants were randomly allocated to a unique pseudonym which was pre-assigned to one of either the Simultaneous fingerprint training condition, Sequential fingerprint training condition, or Control condition. This pseudonym was also used to manage participants' data across sessions.

To commence each session, all participants were required to log in from the experiment home page by selecting their pseudonym from a drop-down box, as well as selecting the appropriate session number (for example, on the first session, participants selected 'Session 1'). Once logged in, participants were asked for demographic information (including gender, age, and whether they had formal experience analysing fingerprints). After entering this information, all participants were then presented with the three xQ tasks. Once the xQ tasks were completed, participants were then automatically directed to their respective training task that was associated with their pseudonym.

2.7.1 Sequential Interleaved Fingerprint Training task. For each trial of the Sequential fingerprint training condition, participants were presented with a single fingerprint at a time on the computer screen. First, participants were required to classify whether the fingerprint came from a left or right hand, by clicking on one of the two category labels presented directly below the fingerprint in question (displayed as ‘Left’, or ‘Right’). Next, for the same fingerprint, participants were asked to then indicate which one of the five finger nominations they believe the print was from (Thumb, Index, Middle, Ring, Little). At the end of each trial, both audio and visual feedback was provided to indicate performance for that trial. For correct responses, the border around the fingerprint would flash green alongside a positive, higher pitched tone. For incorrect responses, the border around the fingerprint would flash red alongside a negative, lower pitched tone. Regardless of the answer the participant selected, the correct answer would appear in the top left corner of the fingerprint alongside the feedback. The order that the fingerprints were presented in was randomised, so as to ensure the categories (e.g., thumb, little, index, etc.) were interleaved. No two fingerprints of the same category appeared consecutively. The training task comprised a total of 100 trials, taking approximately 10 minutes to complete.

2.7.2 Simultaneous Interleaved Training task. For the first trial of the Simultaneous fingerprint training condition, the set-up of the task appeared identical to the Sequential condition. However, for every trial that followed, the fingerprint that had just been answered would move to the top left of the screen, along with the correct category name it belonged to (e.g., Left, Thumb). This print remained in the top left of the screen, while a new print to be classified would appear in the centre of the screen. As more prints were added after each trial, the earlier prints moved across to the right until a total of seven previous prints were displayed across the top of the screen at one time. As the trials continued, the oldest fingerprint would disappear to allow the seven most recent fingerprints to remain. This meant, that for every trial, participants could directly compare the current fingerprint in question with seven other fingerprints that had already been classified. Feedback in this condition was provided in the same way as in the Sequential fingerprint training condition, and the order that fingerprints were presented in was also randomised. The task comprised 100 trials, taking approximately 10 minutes to complete.

2.7.3 Control Word Scramble Task. Participants in the Control condition were presented with seven scrambled letters in the centre of the screen. The task was to select which one of the five unscrambled words to the left matched the scrambled letters. To increase difficulty and maintain stimulation, the five unscrambled words to select an answer from were alphabetically similar to each other. Audio and visual feedback was also provided in this task. The purpose of the word scramble task was to engage a group of participants in a cognitive task that took approximately the same amount of time to complete as the fingerprint training tasks. By having this activity as a control task to make comparisons with, any significant increase in test performance on the xQ that is observed in the fingerprint training conditions, but not the control condition, can be attributed to the fingerprint training tasks specifically.

Once participants had completed all three xQ tasks and their respective training task, this would conclude a session. Participants took part in a total of 11 sessions taking place over 11 consecutive days. However, on the 11th and final session, participants completed the xQ tasks only, as a final test to measure their learning.

Chapter 3: Results

3.1 Descriptive Statistics

Of the 19 participants recruited for this experiment, seven were allocated to the Control condition, six to the Simultaneous fingerprint training condition, and six to the Sequential fingerprint training condition.

Overall, total scores on the xQ tended to be higher for participants in the Simultaneous (*M* = .464, *SD* = .114) and Sequential (*M* = .464, *SD* = .122) fingerprint training conditions in comparison to Controls (*M* = .416, *SD* = .084). However, as can be seen in Figure 1, these differences in scores between conditions do not increase substantially across sessions in any condition. In the Control and Sequential conditions, scores on the xQ at Session 11 tend to show an overall increase from scores on the xQ at Session 1, though it appears that the highest mean scores tend to peak sporadically across sessions, rather than gradually increase to the highest scores at Session 11. Table 1 presents the means and standard deviations of scores on the xQ across sessions for each condition.

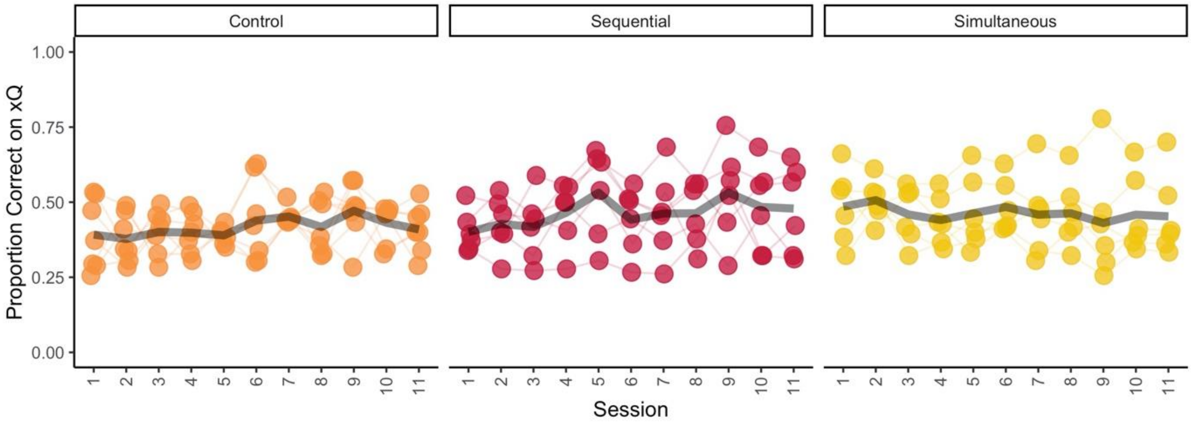


Figure 1. Proportion correct scores on the xQ for individuals in each condition, across sessions

Table 1. Means and standard deviations of xQ scores across sessions, for each condition

Session	Control (SD)	Sequential (SD)	Simultaneous (SD)
1	.392 (.118)	.401 (.163)	.485 (.123)
2	.378 (.081)	.428 (.202)	.506 (.069)
3	.401 (.074)	.418 (.218)	.459 (.095)
4	.398 (.068)	.465 (.201)	.440 (.083)
5	.390 (.028)	.531 (.150)	.463 (.123)
6	.439 (.139)	.442 (.110)	.484 (.089)
7	.452 (.029)	.462 (.143)	.458 (.138)
8	.417 (.091)	.463 (.106)	.463 (.115)
9	.471 (.098)	.531 (.160)	.431 (.187)
10	.432 (.066)	.484 (.145)	.458 (.130)
11	.410 (.080)	.479 (.145)	.453 (.137)

3.2 Confirmatory Analyses

3.2.1 Assumption Checking and Data Screening. Prior to conducting any analyses, I first checked the data for outliers and tested the assumption of normality and sphericity. Visual inspection of box plots revealed there were some outliers in the data. Upon further inspection these outliers were determined to be genuine cases of variation, and thus I did not remove them from subsequent analyses. To test the assumption of normality, the Shapiro-Wilk test of normality was used. Results of this test suggest that scores on the xQ were normally distributed, thus the assumption of normality was met (where $p > .05$ across all combinations of session and training task condition). My within-subjects variable contains 11 levels, therefore I also tested the assumption of sphericity using Mauchly's test of sphericity, which was met, $\chi2(54) = 63.37$, $p = .238$.

3.2.2 Testing Hypothesis 1. It was anticipated that participants in both the Simultaneous and Sequential fingerprint training conditions would gradually perform better on the xQ across sessions compared to participants in the Control condition. To test this hypothesis, I first conducted a mixed factorial ANOVA, to investigate whether there was an interaction effect between training task (between-subjects variable), and session (within-subjects variable). No significant interaction effect between training task and test session was observed, $F(20, 160) = 1.229$, $p = .237$, $\eta^2_G = .062$. I then conducted planned polynomial and treatment contrasts comparing trends in changes on the xQ over sessions between the fingerprint training conditions and the Control condition (Sequential-Control contrast, and Simultaneous-Control contrast). Results suggest that overall, collapsed across sessions, scores on the xQ were significantly higher for both fingerprint training conditions compared to Controls, though these differences were small (Simultaneous: $t(176) = 2.56$, $p = .011$; Sequential: $t(176) = 2.57$, $p = .011$). Trend analysis revealed no significant trend in xQ scores across session for either the Simultaneous ($t = .703$, $p = .483$) or Sequential ($t = -.196$, $p = .845$) fingerprint training conditions (Appendix C).

3.2.3 Testing Hypothesis 2. My second prediction stated that performance improvements on the xQ over sessions would be greater for participants in the Simultaneous compared to the Sequential fingerprint training condition. To test this hypothesis, I conducted polynomial and Helmert contrasts (Appendix D), with the first contrast comparing trends over session between the Sequential fingerprint training condition and the Control condition (which was significant, $\hat{\beta} = .024$, $t = 2.57$, $p = .011$), and the second contrast comparing trends over session between the Simultaneous fingerprint training condition and the Sequential and Control conditions combined. There were no significant differences in scores on the xQ over session between the Simultaneous fingerprint training condition and the Sequential and Control conditions combined ($\hat{\beta} = .007$, $t = 1.436$, $p = 0.153$). However, results of trend analysis revealed a significant negative linear trend in scores across sessions for the Simultaneous condition compared to the Sequential and Controls combined ($\hat{\beta} = -.037$, $t = -2.01$, $p = .046$).

3.3 Post-hoc exploratory analyses

I ran follow up contrasts to explore whether there were significant differences between conditions across sessions in scores on a particular subtask of the xQ, rather than the total xQ score (Appendix E). In particular, I ran contrasts between conditions on the Print Nomination subtask, which most closely resembles the fingerprint training tasks used in this experiment. Results revealed that scores on the Print Nomination subtask increased over sessions significantly more so in the Sequential fingerprint training condition, compared with the Simultaneous and Control conditions combined. This improvement in performance in the Sequential condition was linear ($t = 2.247$, $p = .026$). Figure 2 displays a comparison of performance on the Print Nomination subtask between each condition.

I also ran a mixed ANOVA with the fingerprint training task data as the dependent variable (rather than the xQ) to explore whether there were any significant differences between the two fingerprint training conditions in performance across training sessions, though this result was non-significant, $F(18, 144) = 1.269$, $p = .217$, $\eta^2_G = .068$.

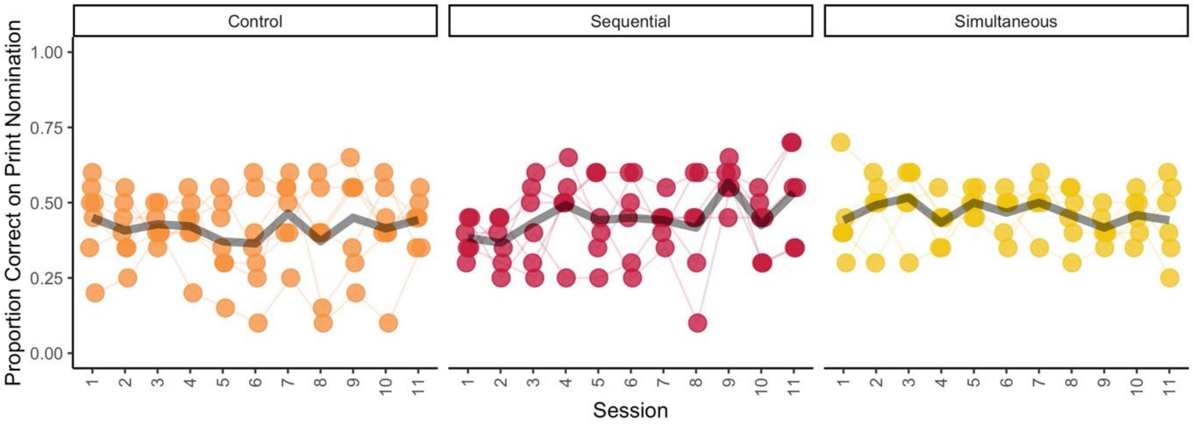


Figure 2. Proportion correct scores on the Print Nomination subtask for individuals in each condition, across sessions.

Chapter 4: Discussion

This study aimed to determine whether a greater learning benefit was produced by presenting stimuli one at a time, or multiple at once, when using interleaved practice as a strategy for studying fingerprints. This experiment drew on well-established learning theories to design fingerprint training tasks that guided learners to compare and contrast the stimuli presented before them, with the overall goal of enhancing their learning. It was anticipated that this study would demonstrate significant differences in learning between conditions that took place gradually across sessions, with differences in learning being the result of either studying fingerprints Simultaneously, Sequentially, or instead participating in a word-scramble task. The results of this experiment will be interpreted in the following sections.

4.1 Addressing Hypothesis 1

It was first anticipated that participants who completed an interleaved fingerprint training task would show greater learning improvements across sessions compared to Controls who completed a word scramble task. The interaction effect between training task and session was not significant, which suggests that any differences in scores on the xQ between training task conditions, were not significantly different across sessions. Additionally, the lack of significant positive trends in scores on the xQ suggests that participants did not demonstrate significant increases in learning across sessions in any condition. Thus, my first hypothesis was not supported. Past research has found that interleaved practice is indeed an effective strategy for use in perceptual learning, particularly when learning categories of stimuli that are highly similar in appearance (Birnbaum et al., 2013; Kang & Pashler, 2012; Zukiply & Burt, 2013). However, the results of this experiment did not demonstrate support for the benefits of interleaved practice to be applicable to the study of fingerprints. The findings in the current study are surprising. Past research that uses other visual categories of stimuli (such as birds or paintings) have found improvements in learning after only a single session of training (Carvalho & Goldstone, 2014a; Kang & Pashler, 2012; Kornell & Bjork, 2008). Yet, after 11 consecutive sessions, this was hardly the case for the current experiment. Possible reasonings to consider for these findings are presented below.

4.1.1 Task-Specificity. A possible explanation for the lack of support for this first hypothesis is that perhaps any learning that did occur from the fingerprint training tasks was not captured by the tasks in the xQ. Perceptual expertise is understood to be highly specific, for example, modern-car enthusiasts cannot transfer their acquired expertise to antique cars (Bukach, Phillips & Gauthier, 2010). Additionally, studies have also shown that holding expertise in a particular area cannot always be demonstrated across different tasks. For example, dog experts have been unable to demonstrate their expertise in a recognition task, nor a matching task, when images of their breed-of-expertise were presented inverted on a computer screen (Robbins & McKone, 2006). Further, humans have an impressive ability to recognise familiar faces, and have even been named ‘familiar face experts’ (Young & Burton, 2018, p.108), however, there is debate in the literature about the nature and extent of this expertise (for contrasting arguments, see Gauthier & Bukach, 2007; McKone & Robbins, 2007). As humans, we are accustomed to seeing faces presented upright through day-to-day encounters, and thus, because this is what we are familiar with, our expertise is limited in that we tend to struggle to identify faces as accurately when they are presented inverted, or upside down (Megreya & Burton, 2006). With reference to the present study, it is possible that any learning resulting from the fingerprint training tasks was very specific to the training task itself, and thus could not be demonstrated on all the tasks in the xQ. As part of an exploratory analysis, I explore this task specificity explanation with regards to one of the individual subtasks in the xQ (which was more alike the training task). This will be discussed under subheading 4.3 (Addressing the Exploratory analyses and the Task-Specificity Explanation).

4.1.2 Desirable difficulties. On a similar note, another possible explanation for why participants in the fingerprint training conditions did not show significantly greater learning across sessions ties in with the theories that underly the concept of desirable difficulties (Bjork, 1994). ‘Desirable difficulties’ is a term that has been given to name conditions of learning that make acquisition more challenging, but in effect can lead to greater long-term retention (Bjork, Soderstrom & Little, 2015). Although seemingly counterintuitive, these conditions of learning tend to produce more errors during learning acquisition in the short term, but they also produce the most learning long term. Interleaved practice has been identified as an example of a desirable difficulty (Bjork & Bjork, 2011), particularly because the to-be-learnt categories are constantly alternating, rather than allowing the learner to study one category at a time. It is evident from the results that participants have made numerous errors across sessions, which shows support for the idea that our interleaved training tasks presented some level of difficulty. Given the nature of our design, we did not measure long term retention, and thus, it cannot be assumed that participants would have demonstrated learning on follow-up tests if they had been administered. However, it is important to consider the possibility that although learning has not been demonstrated through short term performance on the xQ, if long term retention had been measured, it is possible that participants may have then demonstrated significant learning gains.

4.1.3 Overall Group Differences. Interestingly, while participants in the fingerprint training conditions did not show greater learning than Controls over sessions, they did instead perform significantly greater on the xQ than Controls overall, irrespective of session.

Importantly, these differences observed between conditions are small, though nevertheless, results indeed suggest they are significant differences. This finding brings into question then, if these higher scores observed on the xQ are not clearly a result of greater differences in learning that occurred gradually over time, why then did participants who completed the fingerprint training tasks perform better on the xQ overall compared to the Control condition? It is important to note that given the differences between conditions are indeed very small, interpretations of these findings should be considered cautiously. Nevertheless, possible explanations for these significant differences between conditions, particularly in the absence of significant increases across sessions, are discussed below.

4.1.4 Social desirability. Firstly, the data in this experiment were collected under unique circumstances (i.e., within the researcher's home), and comprised a convenience sample of the researcher's friends and family. It is possible that the concept of social desirability, defined as the behaviour of others to "…act in a way consistent with his or her main motivations" (Beauvois & Dubois, 2009, p. 1089) played a role in motivating participants in unique ways. For example, participants in the fingerprint training conditions may have been motivated to perform particularly well so as to achieve more desirable outcomes for the researcher. Contrastingly, for Controls, 11 consecutive days of completing a word scramble task unrelated to fingerprints could be considered enough time to become aware that they were indeed Control participants, and subsequently act accordingly. Taken together, increased motivation for fingerprint training participants to perform well, coupled with reduced motivation for Controls to perform well, may have influenced the significant overall differences that were observed between conditions.

4.1.5 Individual differences. Alternatively, past research has demonstrated that individuals with no prior experience analysing fingerprints tend to have varying levels of baseline ability to accurately differentiate between prints (Searston & Tangen, 2017d). In other words, without any prior study or practice, some individuals tend to perform more accurately than others as a result of individual differences. It is important to consider the role that varying levels of ability among participants may have played in this experiment. It may be the case that as a result of individual differences, some participants achieved higher scores than others overall, though did not necessarily show improvements in scores as the sessions progressed. Given the small sample size in this experiment, and that this small sample was then spread across three conditions, the effect of having even just a small number of higher performing instances in the fingerprint training conditions could be enough to inflate the mean scores of the overall group, particularly in comparison to Controls. This could explain why at the group level, significant differences in performance on the xQ were observed, irrespective of session.

4.2 Addressing Hypothesis 2

The second prediction of this experiment anticipated that participants in the Simultaneous fingerprint training condition would show greater improvements in learning across sessions in comparison to the Sequential fingerprint training condition. Results suggested that the Simultaneous condition did not show significantly greater learning when compared to the Sequential and Control conditions. Interestingly, trend analysis revealed that participants in the Simultaneous condition instead deviated significantly from participants in the Sequential condition in a downwards linear trend. This suggests that across sessions, the Simultaneous condition instead performed worse than the Sequential and Control condition. Therefore, the second hypothesis of this experiment was not supported. These results were also surprising, as past research using other perceptual categories of stimuli have consistently demonstrated greater learning in participants who study using simultaneous sequences of interleaved practice in comparison to sequential sequences (Carvalho & Goldstone, 2014a; Wahlheim et al., 2011).

4.2.1 Limitations of side-by-side comparisons. It is understood that simultaneous presentations of interleaved stimuli are most beneficial for learners because this allows the differences between categories to be directly compared side-by-side (See Carvalho & Goldstone, 2014a). Not only do simultaneous presentations make the differences between categories more salient, but it also reduces the strain on working memory to retain the intricate details of each stimulus. This reduced strain on memory means that more effort can be put towards comparing the side-by-side stimuli, increasing the ability to correctly notice whether specific features differ. However, perhaps what has been an assumed benefit of making these direct comparisons has instead played a disadvantage for the Simultaneous condition of this experiment. By allowing participants to compare stimuli side-by-side, there was less pressure for these participants to retain the details of previous fingerprints that differed between categories, and therefore, less of this information was being actively processed. In effect, this has inadvertently taken away the element of desirable difficulties in the Simultaneous condition that this study had taken advantage of, by making it easier to make comparisons between categories. On the other hand, the Sequential condition viewed fingerprint stimuli one at a time, and thus, the strain on memory to accurately remember the details of each stimulus remained. Though it may have been ‘easier’ for the Simultaneous condition to compare and contrast stimuli side-by-side, the continuous retrieval of information that is important for learning (Endres, Carpenter, Martin & Renkl, 2017) was likely not as utilised in the Simultaneous condition, as it was in the Sequential condition. Therefore, this continuous retrieval process may have benefitted the Sequential condition, but not the Simultaneous condition, and in effect this enhanced learning in the Sequential condition instead. This provides a possible explanation for why the Simultaneous condition of this experiment did not perform significantly greater than the Sequential condition, and instead performed worse across sessions in comparison.

4.3 Addressing the Exploratory Analyses and the Task-Specificity Explanation

Taking into consideration the task-specificity explanation that was discussed earlier for the lack of support for the first hypothesis (4.1.1 Task-Specificity), it seemed appropriate to explore participants' scores on the Print Nomination subtask of the xQ. The Print Nomination task required participants to identify whether each fingerprint in question came from a left or right hand, and whether the print was also from a thumb, index, middle, ring, or little finger. This task was mostly similar to the fingerprint training tasks used in this experiment, with the only differences in the training tasks being the addition of immediate feedback, a greater number of trials, and the ability to compare previous fingerprints side-by-side (in the Simultaneous condition). Thus, given the similarity in tasks, the task-specificity explanation would suggest that participants who completed the fingerprint training tasks should have been able to demonstrate improved performance across sessions specifically on the Print Nomination task. Interestingly, results of this exploratory analysis showed partial support for this idea: Participants in the Sequential condition showed greater significant improvements across sessions compared to the Simultaneous and Control conditions, while the Simultaneous condition did not show any significant improvements across sessions at all. The task-specificity explanation might expect that the Simultaneous condition should have also significantly improved across sessions, given they also completed the fingerprint training tasks. However, taken together with the possibility that perhaps the side-by-side comparisons in the Simultaneous condition is actually a potential disadvantage for learning, the findings of these post-hoc analyses may indeed be logical.

4.4 Limitations

Like with any experiment, this study is not without its limitations. Firstly, the conditions under which data were collected could not be controlled in the same way they could be in standard laboratory settings. Participants were not strangers to each other, and though they were requested not to discuss the tasks involved in the experiment until all sessions had concluded, this is not something that could always be easily monitored. Experiment-related communication may have taken place between participants of different conditions by accident, which could have impacted participants' motivations if their conditions were to be accidentally revealed (i.e., experimental versus control conditions).

Secondly, though this study did collect longitudinal data, it was limited to the 11 consecutive days of the experiment and did not measure a long-term retention element of learning. In line with the notion of desirable difficulties (Bjork, 1994), exposing participants to study material that is difficult for the learner - as has been done in this experiment - is likely to produce greater learning outcomes long-term, which may not be demonstrated in the short term. Had this study implemented additional points of data collection following a period of delay, perhaps deeper insight into participants' learning could have been revealed.

Lastly, participants that make up the sample in this experiment are not an accurate representation of the true population of novices. With only 19 participants split across three conditions, this is difficult to achieve, though these 19 participants were also drawn from the inner circles of the researchers. Naturally, this places limits on the participant pool in terms of variation in individual factors that would be true of a wider population of novices in reality. Friends and family members are likely to have several characteristics in common (genetics, socio-economic status, education, etc.) and this lack of diversity in the sample makes these results difficult to generalise to a wider population.

4.5 Strengths

Despite these limitations, this experiment holds considerable strengths that are worth noting. Firstly, rather than adopting a pre-test/post-test-only design, this experiment instead captured data across each of the 11 Sessions. This has demonstrated its benefits when taking into consideration the nature of the trends in the data. In two out of three conditions of this experiment, average scores at the group level show that final scores at Session 11 (which would equate to a post-test) have increased compared to initial scores at Session 1 (which would equate to a pre-test). If this pre/post-test data was all that had been captured, it may have provided a false sense of success in the fingerprint training conditions, where all we would see is an increase in scores. However, as this experiment captured data across each session, it is possible to visualise that this increase from Session 11 compared to Session 1 is not so straightforward, and there are a number of peaks and troughs in scores over the duration of the experiment.

Secondly, the order in which the tasks of this experiment were presented to participants is also a notable strength. When testing materials are presented immediately after learning and instruction, often what is captured is an individual's temporarily inflated performance, though this is not a reflection of their permanently stored knowledge (Bjork, 1994; Soderstrom & Bjork, 2015). With regards to this experiment, participants always completed the xQ measurement of fingerprint expertise first, which was then followed by their respective training tasks. By presenting our measurement first, then followed by the training tasks, this experiment has avoided capturing any artificial inflations in test scores that may have arisen if materials were presented in the reverse order.

4.6 Future Directions

This experiment has paved the way for future studies that may wish to further explore interleaved practice in fingerprint stimuli, though perhaps with the inclusion of some important differences. For example, distributing training sessions further across time, such as once a week or once a month, rather than once a day. Distributing training sessions across time has been beneficial in the study of other stimuli (such as Spanish words). For example, participants who completed training every 30 days for three sessions, showed greater long-term retention compared to participants who studied three sessions over three consecutive days (Bahrick, 1979). In reality, trainee fingerprint examiners undertake approximately five years of training before becoming qualified (Searston & Tangen, 2017d), so perhaps 11 consecutive days of training is not enough to capture the first glimpses of improvements in learning. Thus, researchers may choose to explore whether extended timelines of training sessions, potentially with larger delays between each session, may benefit the study of fingerprints.

Employing a variety of different tasks in a training protocol may also be a useful consideration for future research. Introducing variation in training can enhance a learner's ability to transfer acquired skills across to new (but relevant) environments (Bjork, 1994). Therefore, researchers may choose to include a mixture of different tasks in training, such as a visual search task (Searston & Tangen, 2017a), or a simple fingerprint matching task, in order to increase the likelihood of participants being able to demonstrate their learnt skills across a variety of tasks.

Lastly, research in the education domain has found that testing participants using short answer responses with immediate feedback produces greater learning than multiple choice quizzes (Kang, McDermott & Roedigger III, 2007). Further, asking individuals to elaborate on a response by providing an explanation for an answer has also been shown to enhance recall

(Pressley, McDaniel, Turnure, Wood & Ahmad, 1987). Thus, including a short answer element to interleaved training tasks that prompts participants to explain why they have selected a certain answer may be an avenue to explore in future research.

4.7 Implications

It is important to consider the following implications with respect to the limitations that have been discussed. Nevertheless, this research provides important theoretical and practical implications for both the perceptual learning domain as well as the broader forensic field of fingerprint examination.

From a theoretical perspective, this research has expanded on the literature surrounding interleaved practice by using a novel paradigm that allowed for Simultaneous and Sequential presentations of interleaved fingerprint stimuli to be compared. The findings of the current experiment have potentially identified instances where Sequential presentations of interleaved stimuli are more beneficial than Simultaneous presentations. It has previously been understood that Simultaneous presentations are more beneficial for learning, particularly when the stimuli being learned are highly similar in appearance, which undoubtedly describes our fingerprint stimuli (Birnbaum et al., 2013; Carvalho & Goldstone, 2014b). Our findings call for the need to reconsider the types of stimuli that will most benefit from different presentations of interleaved practice, and the mechanisms behind what makes each presentation more or less beneficial. It may be the case that in some instances, as has been found in this experiment, Simultaneous presentations of interleaved stimuli serve to benefit the learner’s ability to compare those stimuli, but do not necessarily facilitate or enhance learning.

This research also has important practical implications. In the field of fingerprint examination, ensuring accuracy in decision-making for both established veterans and new-to-the-field trainees alike, is at the forefront of priorities. Developing empirically based training tools for new recruits is an important step required to achieve these goals of making minimal errors in practice for future generations of forensic professionals. This experiment has undertaken exploration of one possible learning strategy with different presentation styles that could be considered when designing future training programs for novices. The results provide crucial insight into the effectiveness of interleaved practice as a learning strategy in fingerprints, with particular focus on how best to present stimuli within interleaved practice. With additional work, researchers can become more confident in what strategies will demonstrate the most advantages for learning in the fingerprint domain. In itself, exploring how best to effectively train new fingerprint examination recruits has a realm of beneficial follow-on effects, including improved forensic decision-making in those with less experience, and reduced miscarriages of justice.

4.8 Concluding Remarks

This experiment aimed to investigate whether a greater learning benefit was produced among novice participants when interleaved fingerprint stimuli were presented one at a time (Sequentially), or multiple at one time (Simultaneously). Using novel computer-based fingerprint training tasks, participants completed 11 consecutive days of testing and training alongside Control participants who completed a word scramble task. There were indeed some interesting findings in this experiment, such as instances of the Sequential condition performing better than the Simultaneous condition, which was not initially anticipated. However, overall, any significant improvements in learning across sessions were indeed very small.

What has become apparent in the current experiment, is that what seems promising in theory is not always easily transferred when applying it to practice. For example, where past research suggests that there are benefits to using interleaved practice in the study of birds or paintings, there are many factors that need to be considered before anticipating that these same benefits will be observable in the applied domain of fingerprint training. In hindsight, 11 consecutive days of a single task in fingerprint training seems incomparable to the five years of training required in real life practice. Thus, future studies should consider distributing training sessions further across time, as well as incorporate a mixture of different tasks during fingerprint training, in an attempt to maximise learning in novice individuals. Despite the lack of support for the hypotheses in this experiment, this study provides important contributions that bring researchers one step closer to identifying the most optimal way to train fingerprint examiners of the future.

References

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology*, *108*(3), 296-308.

Beauvois, J. L., & Dubois, N. (2009). Lay psychology and the social value of persons. *Social and Personality Psychology Compass, 6*(3), 1082-1095.

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory and Cognition, 41*(3)*,* 392-402.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about Knowing* (pp.185-205). Cambridge, MA: Massachusetts Institute of Technology Press

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough., & J. R. Pomerantz. (Eds). *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society* (pp. 56-64). New York: Worth Publishers.

Bjork, E. L., Soderstrom, N. C., & Little, J. L. (2015). Can multiple-choice testing induce desirable difficulties? Evidence from the laboratory and the classroom. *The American Journal of Psychology, 128*(2), 229-239.

Bukach, C. M., Phillips, W. S., & Gauthier, I. (2010). Limits of generalization between categories and implications for theories of category specificity. *Attention, Perception and Psychophysics, 72*(7), 1865-74.

Busey, T. A., Parada, F. J. (2010). The nature of expertise in fingerprint examiners. *Psychonomic Bulletin and Review, 17*(2), 155-160.

Carvalho, P. F., & Goldstone, R. L. (2014a). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory and Cognition, 42*(3), 481-495.

Carvalho, P. F., & Goldstone (2014b). Effects of interleaved and blocked study on delayed test of category learning generalization. *Frontiers in Psychology, 5*(936), 1-11.

Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction, 49*, 13-20.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*(1), 273-305.

Finley, J. R., Roedigger III, H. L., Hughes, A. D., Wahlheim, C. N., & Jacoby, L. L. (2015). Simultaneous versus sequential presentation in testing recognition memory for faces. *The American Journal of Psychology, 128*(2), 173-195.

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition, 103*(2), 322-330.

Kang, S. H. K., McDermott, K. B., & Roedigger III, H. L. (2017). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*(4), 528-558.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing in advantageous when it comes to discriminative contrast. *Applied Cognitive Psychology, 26*(1), 97-103.

Kellman, P. J., & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews, 6*(2), 53-84.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “Enemy of Induction”? *Psychological Science, 19*(6), 585-592.

Martire, K. A., & Edmond, G. (2017). Rethinking expert opinion evidence. *Melbourne University Law Review, 40*(3), 967-998.

McKone, E., & Robbins, R. (2007). The evidence rejects the expertise hypothesis: Reply to Gauthier & Bukach. *Cognition, 103*(2), 331-336.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207-238.

Megraya, A. M., & Burton, (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory and Cognition, 34*(4), 865-876.

Michalski, R. S. (1983). A theory and methodology of inductive learning. *Artificial Intelligence, 20*(2), 111-161.

Mnookin, J. L. (2008). Expert evidence, partisanship, and epistemic competence. *Brooklyn Law Review, 73*(3), 1009-1033.

Nosofsky, R. (2011). The generalized context model: an exemplar model of classification. In E. M. Pothos, & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 18–39). New York: Cambridge University Press.

Pressly, M., McDaniel, M. A., Turnure, J. E., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects of intentional and incidental learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13*(2), 291-300.

Rieber, M. (1966). Role of stimulus comparison in children's discrimination learning. *Journal of Experimental Psychology, 72*(2), 263-270.

Robbins, R., & McKone, E. (2006). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition, 103*(1), 34-79.

Searston, R. A., & Chin, J. M. (2019). The legal and scientific challenge of black box expertise. *University of Queensland Law Journal, 38*(2), 237-260.

Searston, R. A., & Tangen, J. M. (2017a). Expertise with unfamiliar objects is flexible to changes in task but not changes in class. *PLoS ONE, 12*(6), 1-14.

Searston, R. A., & Tangen, J. M. (2017b). The style of a stranger: Identification expertise generalizes to coarser level categories. *Psychonomic Bulletin Review, 24*(4), 1324-1329.

Searston, R., & Tangen, J. (2017c). Training perceptual experts: Feedback, labels, and contrasts. *Canadian Journal of Experimental Psychology*, *71*(1), 32-39.

Searston, R. A., & Tangen, J. M. (2017d). The emergence of perceptual expertise with fingerprints over time. *Journal of Applied Research in Memory and Cognition, 6*(4), 442-451.

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10*(2), 176-199.

Stevenage, S. V., & Pitfield, C. (2016). Data from a fingerprint matching task with experts, trained students and untrained novices. *Data in Brief, 9*, 621-624.

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science, 16*(2), 145-151.

Tangen, J. M. (2013). Identification personified. *Australian Journal of Forensic Sciences, 45*(3), 315-322.

Tangen, J. M., Thompson, M. B., & McCarthy, D. J. (2011). Identifying fingerprint expertise. *Psychological Science, 22*(8), 995-997.

Thompson, M. B., & Tangen, J. M. (2014). The nature of expertise in fingerprint matching: Experts can do a lot with a little. *PLoS ONE, 9*(12), 1-23.

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2013). Expertise in fingerprint identification. *Forensic Sciences, 58*(6), 1519-1530.

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2014). Human matching performance of genuine crime scene latent fingerprints. *Law and Human Behaviour, 38*(1), 84-93.

Towler, A., White, D., Ballantyne, K., Searston, R. A., Martire, K. A., & Kemp, R. I. (2018). Are forensic scientists experts? *Journal of Applied Research in Memory and Cognition, 7*(2)*,* 199-208.

Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2011). Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences, 108*(19), 7733-7738.

Vojdanoska, M., Cranney, J., & Newell, B. R. (2010). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology, 24*(8), 1183-1195.

Vokey, J. R., Tangen, J. M., & Cole, S. A. (2009). On the preliminary psychophysics of fingerprint identification. *The Quarterly Journal of Experimental Psychology, 62*(5), 1023-1040.

Wahlheim, C. N., Dunlosky, J., Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory and Cognition, 39*(5)*,* 750-763.

Westfall, J. (2020). PANGEA: Power analysis for general ANOVA designs. Retrieved from: https://jakewestfall.shinyapps.io/pangea/

Yan, V. X., Soderstrom, N. C., Seneviratna, G. S., Bjork, L. L., Bjork, R. A. (2017). How should exemplars be sequenced in inductive learning? Empirical evidence versus learners' opinions. *Journal of Experimental Psychology: Applied, 23*(4), 403-416.

Young, A. W., & Burton, A. M. (2018). Are we face experts? *Trends in Cognitive Science, 22*(2), 100-110.

Zukiply, N., Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory and Cognition, 41*(1), 16-27.

Appendix A: Participant Information Sheet

Participant Information

THE UNIVERSITY of ADELAIDE

PROJECT: Examining the effect of simultaneous versus sequential interleaved learning on categorising fingerprints
RESEARCHER: Jade Cascun, School of Psychology, University of Adelaide
SUPERVISOR: Dr Rachel Searston, School of Psychology, University of Adelaide
LOCATION: Adelaide, South Australia and Melbourne, Victoria.

PARTICIPATION: You are invited to participate in a study of human perception and cognition. We hope to understand whether certain methods of learning can be considered more advantageous than others within the domain of fingerprints. This study involves participating in a learning task which requires you to make judgements about stimuli and classifying them into the categories you believe are correct. The results from this experiment could potentially be useful in working towards the larger goals of forensic disciplines who are aiming to enhance the quality of training for expert forensic examiners, and provide an empirical background for such training. There are no foreseeable risks associated with this study that are any different to those of daily life. This study will take place over 11 sessions for 11 consecutive days, with each session taking approximately 20 minutes (with the exclusion of the first day, where you will be required to undertake a simple baseline performance test in addition to the daily training session). The researcher will be available nearby throughout all sessions. After you have completed the experiment, the researcher will discuss with you the overall study, explain the methodology of the experiment, the variables of interest, and answer any questions you have.

CONSENT: Participation in this study is entirely voluntary. You are free to withdraw from the study at any time and will not be penalized in any way. You will receive a $20 Coles/Myer gift card for every hour of participation in the experiment, and will be asked to sign an acknowledgment of payment form. If, for any reason, you do not want to continue with the experiment, simply let the researcher know. In this event you will still be awarded full payment.

DATA MANAGEMENT: Any information that is obtained from this study will remain entirely confidential and will be kept on a password protected computer with multiple redundant backups. The data from this experiment will be identified via use of a pseudonym in order to protect participant privacy. You will not be identifiable by this pseudonym, but your performance in this experiment will be recorded (though not associated with you personally). We plan to discuss the results at academic conferences both here and overseas, publish the data in international scientific journals, and store the data in an online open access repository, such as the Open Science Framework, for future meta-analyses and so that other researchers can easily reproduce our work. In any publication, presentation or online record, you cannot be identified.

ETHICS: The study has been approved by the School of Psychology Human Research Ethics Subcommittee at the University of Adelaide (approval number: 20/29). This research will be conducted according to the NHMRC National Statement on Ethical Conduct in Human Research (2007). For any questions about the ethical conduct of the research, please contact Professor Paul Delfabbro, Chair of the Human Research Ethics Subcommittee in the School of Psychology (paul.delfabbro@adelaide.edu.au).

Rachel Searston, PhD
School of Psychology
The University of Adelaide
rachel.searston@adelaide.edu.au

Project Number: 20/29

Appendix B: Consent Form

Consent Form

THE UNIVERSITY *of* ADELAIDE

PROJECT: Examining the effect of simultaneous versus sequential interleaved learning on categorising fingerprints
RESEARCHER: Jade Cascun, School of Psychology, University of Adelaide
SUPERVISOR: Dr Rachel Searston, School of Psychology, University of Adelaide
LOCATION: Adelaide, South Australia and Melbourne, Victoria.

1. I agree to participate in the project named above, which is for research purposes. The particulars of the project, including details of the tasks, have been explained to me and provided to me via the participant information sheet.

2. I consent to any data gathered from this participation to be used for research purposes and to the data being uploaded and stored in an online public repository (e.g., Open Science Framework), available to other researchers.

3. I consent to any data gathered from this participation to be presented to non-academic bodies (e.g., fingerprint examiners) if the research is deemed to be useful to their discipline.

4. I acknowledge that:
    (a) the project is for the purpose of research;
    (b) I have been informed that my involvement is voluntary and that I am free to withdraw from the project at any time without explanation or prejudice and to withdraw any unprocessed data previously supplied;
    (c) the possible effects of the tasks have been explained to me to my satisfaction; and
    (d) I have been informed that the confidentiality of the information I provide will be protected subject to any legal requirements.

5. I understand that:
    (a) my real name or any other identifiable data will not be used in any publications arising from the research without my consent; and

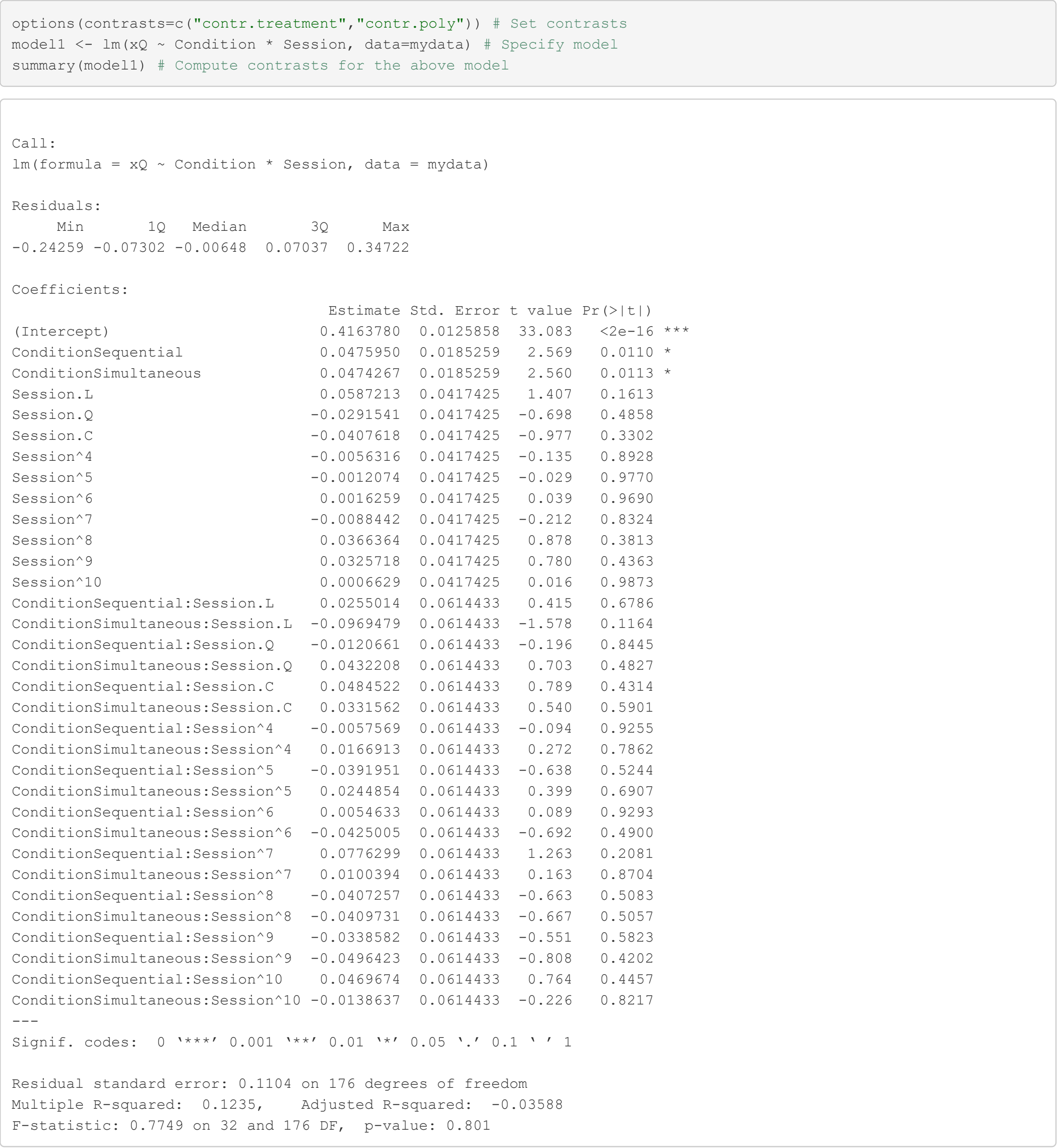
Name of participant: ____

Signature: ______________________________ Date: ______________

Rachel Searston, PhD
School of Psychology
The University of Adelaide
rachel.searston@adelaide.edu.au

Project Number: 20/29

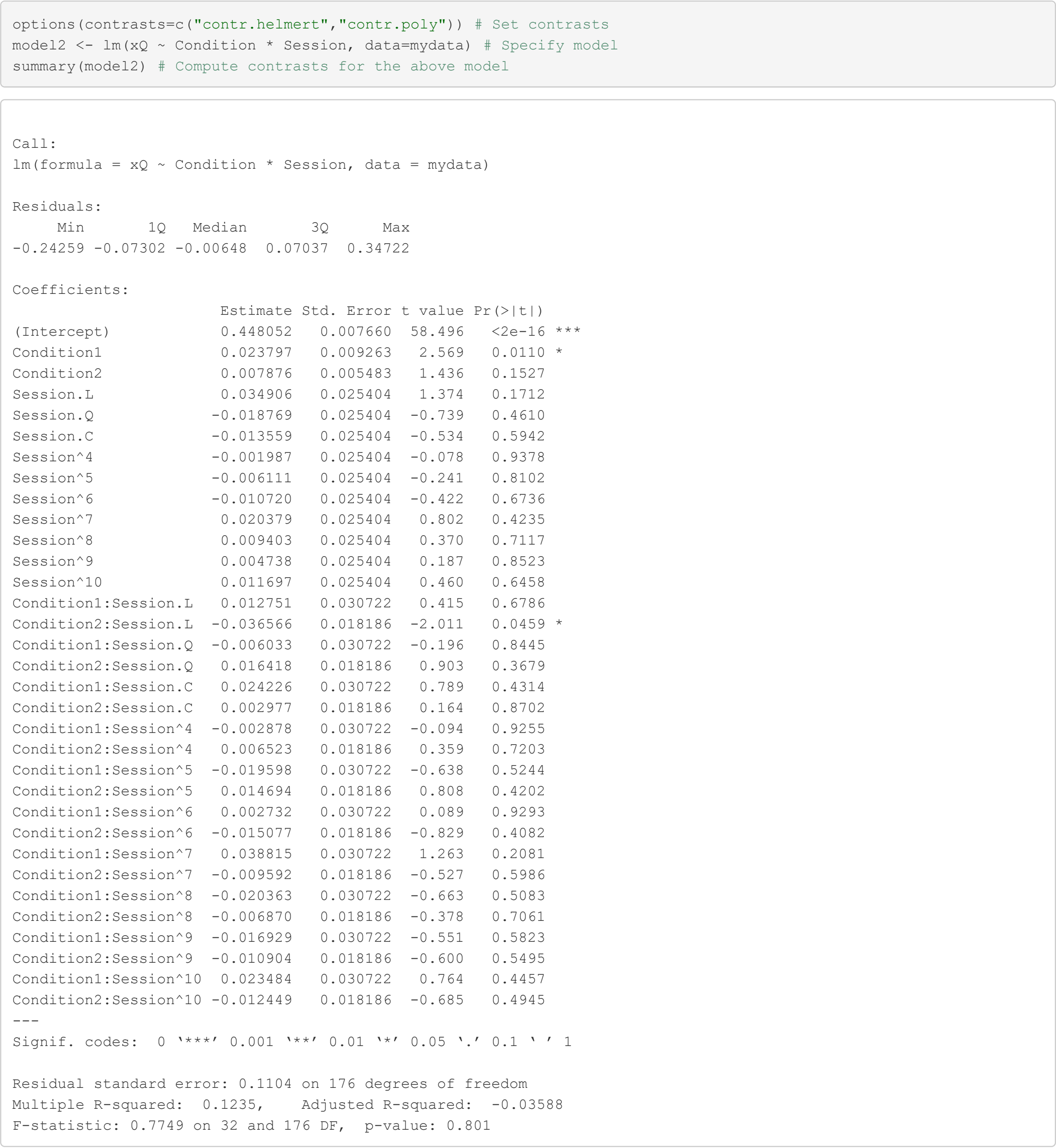
Appendix C: R code and output for treatment contrasts and trend analysis



```
options(contrasts=c("contr.treatment","contr.poly")) # Set contrasts
model1 <- lm(xQ ~ Condition * Session, data=mydata) # Specify model
summary(model1) # Compute contrasts for the above model
```

```

Call:
lm(formula = xQ ~ Condition * Session, data = mydata)

Residuals:
     Min       1Q   Median       3Q      Max
-0.24259 -0.07302 -0.00648  0.07037  0.34722

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                       0.4163780  0.0125858  33.083   <2e-16 ***
ConditionSequential               0.0475950  0.0185259   2.569   0.0110 *
ConditionSimultaneous             0.0474267  0.0185259   2.560   0.0113 *
Session.L                         0.0587213  0.0417425   1.407   0.1613
Session.Q                        -0.0291541  0.0417425  -0.698   0.4858
Session.C                        -0.0407618  0.0417425  -0.977   0.3302
Session^4                        -0.0056316  0.0417425  -0.135   0.8928
Session^5                        -0.0012074  0.0417425  -0.029   0.9770
Session^6                         0.0016259  0.0417425   0.039   0.9690
Session^7                        -0.0088442  0.0417425  -0.212   0.8324
Session^8                         0.0366364  0.0417425   0.878   0.3813
Session^9                         0.0325718  0.0417425   0.780   0.4363
Session^10                        0.0006629  0.0417425   0.016   0.9873
ConditionSequential:Session.L     0.0255014  0.0614433   0.415   0.6786
ConditionSimultaneous:Session.L  -0.0969479  0.0614433  -1.578   0.1164
ConditionSequential:Session.Q    -0.0120661  0.0614433  -0.196   0.8445
ConditionSimultaneous:Session.Q   0.0432208  0.0614433   0.703   0.4827
ConditionSequential:Session.C     0.0484522  0.0614433   0.789   0.4314
ConditionSimultaneous:Session.C   0.0331562  0.0614433   0.540   0.5901
ConditionSequential:Session^4    -0.0057569  0.0614433  -0.094   0.9255
ConditionSimultaneous:Session^4   0.0166913  0.0614433   0.272   0.7862
ConditionSequential:Session^5    -0.0391951  0.0614433  -0.638   0.5244
ConditionSimultaneous:Session^5   0.0244854  0.0614433   0.399   0.6907
ConditionSequential:Session^6     0.0054633  0.0614433   0.089   0.9293
ConditionSimultaneous:Session^6  -0.0425005  0.0614433  -0.692   0.4900
ConditionSequential:Session^7     0.0776299  0.0614433   1.263   0.2081
ConditionSimultaneous:Session^7   0.0100394  0.0614433   0.163   0.8704
ConditionSequential:Session^8    -0.0407257  0.0614433  -0.663   0.5083
ConditionSimultaneous:Session^8  -0.0409731  0.0614433  -0.667   0.5057
ConditionSequential:Session^9    -0.0338582  0.0614433  -0.551   0.5823
ConditionSimultaneous:Session^9  -0.0496423  0.0614433  -0.808   0.4202
ConditionSequential:Session^10    0.0469674  0.0614433   0.764   0.4457
ConditionSimultaneous:Session^10 -0.0138637  0.0614433  -0.226   0.8217
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1104 on 176 degrees of freedom
Multiple R-squared:  0.1235,    Adjusted R-squared:  -0.03588
F-statistic: 0.7749 on 32 and 176 DF,  p-value: 0.801
```

Appendix D: R code and output for Helmert contrasts, and trend analysis on xQ scores between conditions across sessions

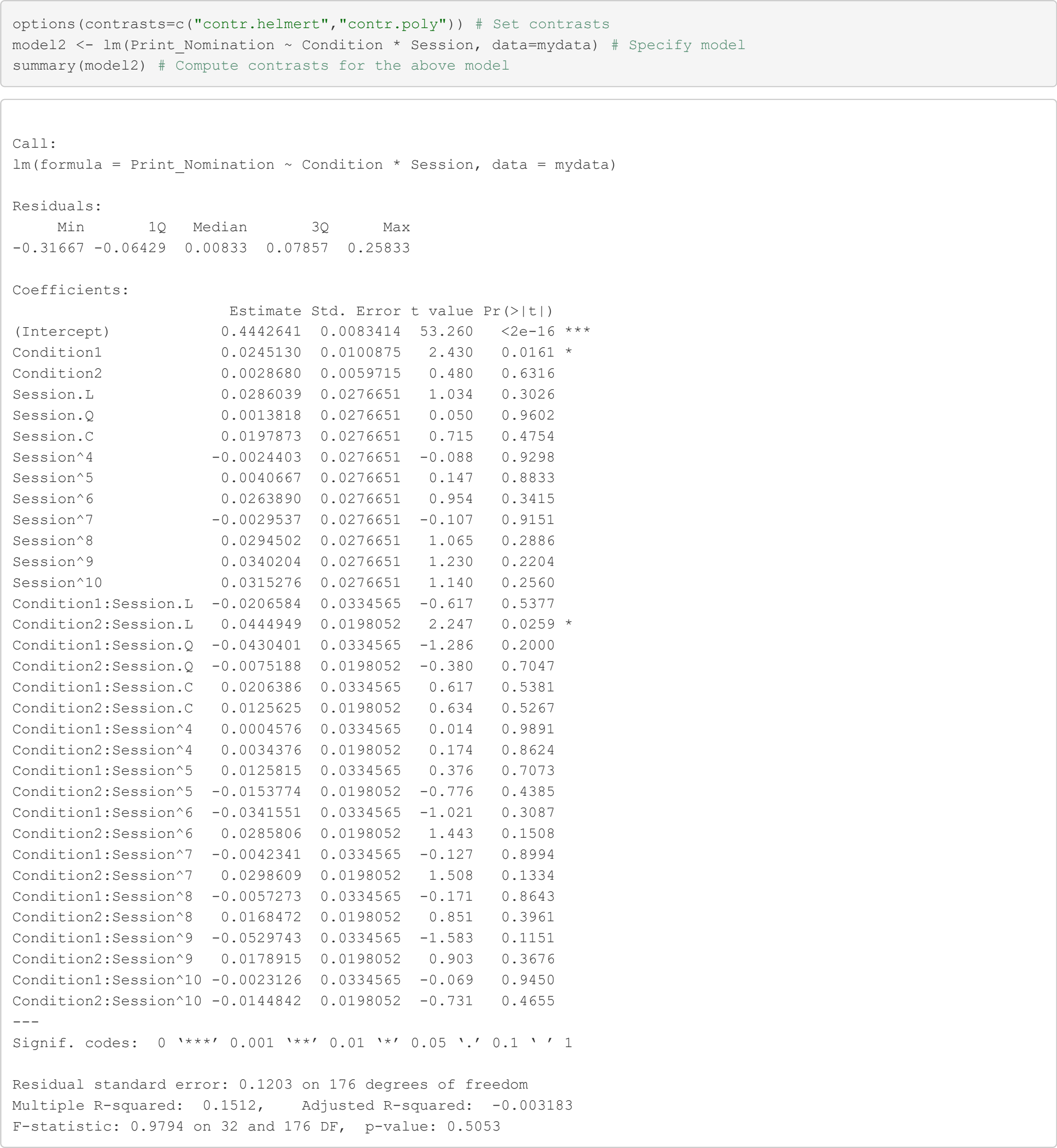


```r
options(contrasts=c("contr.helmert","contr.poly")) # Set contrasts
model2 <- lm(xQ ~ Condition * Session, data=mydata) # Specify model
summary(model2) # Compute contrasts for the above model
```

```
Call:
lm(formula = xQ ~ Condition * Session, data = mydata)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.24259 -0.07302 -0.00648  0.07037  0.34722 

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)            0.448052   0.007660  58.496   <2e-16 ***
Condition1             0.023797   0.009263   2.569   0.0110 *  
Condition2             0.007876   0.005483   1.436   0.1527    
Session.L              0.034906   0.025404   1.374   0.1712    
Session.Q             -0.018769   0.025404  -0.739   0.4610    
Session.C             -0.013559   0.025404  -0.534   0.5942    
Session^4             -0.001987   0.025404  -0.078   0.9378    
Session^5             -0.006111   0.025404  -0.241   0.8102    
Session^6             -0.010720   0.025404  -0.422   0.6736    
Session^7              0.020379   0.025404   0.802   0.4235    
Session^8              0.009403   0.025404   0.370   0.7117    
Session^9              0.004738   0.025404   0.187   0.8523    
Session^10             0.011697   0.025404   0.460   0.6458    
Condition1:Session.L   0.012751   0.030722   0.415   0.6786    
Condition2:Session.L  -0.036566   0.018186  -2.011   0.0459 *  
Condition1:Session.Q  -0.006033   0.030722  -0.196   0.8445    
Condition2:Session.Q   0.016418   0.018186   0.903   0.3679    
Condition1:Session.C   0.024226   0.030722   0.789   0.4314    
Condition2:Session.C   0.002977   0.018186   0.164   0.8702    
Condition1:Session^4  -0.002878   0.030722  -0.094   0.9255    
Condition2:Session^4   0.006523   0.018186   0.359   0.7203    
Condition1:Session^5  -0.019598   0.030722  -0.638   0.5244    
Condition2:Session^5   0.014694   0.018186   0.808   0.4202    
Condition1:Session^6   0.002732   0.030722   0.089   0.9293    
Condition2:Session^6  -0.015077   0.018186  -0.829   0.4082    
Condition1:Session^7   0.038815   0.030722   1.263   0.2081    
Condition2:Session^7  -0.009592   0.018186  -0.527   0.5986    
Condition1:Session^8  -0.020363   0.030722  -0.663   0.5083    
Condition2:Session^8  -0.006870   0.018186  -0.378   0.7061    
Condition1:Session^9  -0.016929   0.030722  -0.551   0.5823    
Condition2:Session^9  -0.010904   0.018186  -0.600   0.5495    
Condition1:Session^10  0.023484   0.030722   0.764   0.4457    
Condition2:Session^10 -0.012449   0.018186  -0.685   0.4945    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1104 on 176 degrees of freedom
Multiple R-squared:  0.1235,	Adjusted R-squared:  -0.03588 
F-statistic: 0.7749 on 32 and 176 DF,  p-value: 0.801
```

Note: Condition 1 = Sequential - Control contrast. Condition 2 = Simultaneous compared to Sequential-Control combined contrast.

Appendix E: Exploratory Helmert contrasts and trend analysis on Print Nomination subtask, between conditions across sessions



```r
options(contrasts=c("contr.helmert","contr.poly")) # Set contrasts
model2 <- lm(Print_Nomination ~ Condition * Session, data=mydata) # Specify model
summary(model2) # Compute contrasts for the above model
```

```
Call:
lm(formula = Print_Nomination ~ Condition * Session, data = mydata)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.31667 -0.06429  0.00833  0.07857  0.25833 

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)            0.4442641  0.0083414  53.260   <2e-16 ***
Condition1             0.0245130  0.0100875   2.430   0.0161 *  
Condition2             0.0028680  0.0059715   0.480   0.6316    
Session.L              0.0286039  0.0276651   1.034   0.3026    
Session.Q              0.0013818  0.0276651   0.050   0.9602    
Session.C              0.0197873  0.0276651   0.715   0.4754    
Session^4             -0.0024403  0.0276651  -0.088   0.9298    
Session^5              0.0040667  0.0276651   0.147   0.8833    
Session^6              0.0263890  0.0276651   0.954   0.3415    
Session^7             -0.0029537  0.0276651  -0.107   0.9151    
Session^8              0.0294502  0.0276651   1.065   0.2886    
Session^9              0.0340204  0.0276651   1.230   0.2204    
Session^10             0.0315276  0.0276651   1.140   0.2560    
Condition1:Session.L  -0.0206584  0.0334565  -0.617   0.5377    
Condition2:Session.L   0.0444949  0.0198052   2.247   0.0259 *  
Condition1:Session.Q  -0.0430401  0.0334565  -1.286   0.2000    
Condition2:Session.Q  -0.0075188  0.0198052  -0.380   0.7047    
Condition1:Session.C   0.0206386  0.0334565   0.617   0.5381    
Condition2:Session.C   0.0125625  0.0198052   0.634   0.5267    
Condition1:Session^4   0.0004576  0.0334565   0.014   0.9891    
Condition2:Session^4   0.0034376  0.0198052   0.174   0.8624    
Condition1:Session^5   0.0125815  0.0334565   0.376   0.7073    
Condition2:Session^5  -0.0153774  0.0198052  -0.776   0.4385    
Condition1:Session^6  -0.0341551  0.0334565  -1.021   0.3087    
Condition2:Session^6   0.0285806  0.0198052   1.443   0.1508    
Condition1:Session^7  -0.0042341  0.0334565  -0.127   0.8994    
Condition2:Session^7   0.0298609  0.0198052   1.508   0.1334    
Condition1:Session^8  -0.0057273  0.0334565  -0.171   0.8643    
Condition2:Session^8   0.0168472  0.0198052   0.851   0.3961    
Condition1:Session^9  -0.0529743  0.0334565  -1.583   0.1151    
Condition2:Session^9   0.0178915  0.0198052   0.903   0.3676    
Condition1:Session^10 -0.0023126  0.0334565  -0.069   0.9450    
Condition2:Session^10 -0.0144842  0.0198052  -0.731   0.4655    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1203 on 176 degrees of freedom
Multiple R-squared:  0.1512,	Adjusted R-squared:  -0.003183 
F-statistic: 0.9794 on 32 and 176 DF,  p-value: 0.5053
```

Note: Condition 1 = Simultaneous – Controls contrast. Condition 2 = Sequential and Simultaneous – Controls combined contrast