Mixing it Up: Developing Expertise in Forensic Fingerprint Examination Using Interleaved Practice

Claire Walker

University of Adelaide

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

Submitted on 29 September, 2020

Word Count: 9,478

Table of Contents

List of Figures

List of Tables

Abstract

Forensic fingerprint experts have a superior ability to differentiate highly similar print pairs, especially in comparison to novices (those with no experience in the interpretation of fingerprints). Few studies have investigated methods of effectively training novices to become experts. The current study draws on the principle of interleaved practice to train a small sample of fingerprint novices. Interleaving theory purports that 'mixing' exemplars from different categories has greater learning benefit than 'massing' exemplars from the same category. The current experiment applied this principle via a novel training paradigm in which one group of novices responded to fingerprints from different fingers (Mixed), and a second group responded to fingerprints from the same finger (Massed). An active control group completed a task unrelated to fingerprint examination. All participants completed a measure of fingerprint expertise performance (the xQ) immediately prior to each of 10 training sessions across 10 consecutive days, with a final measure of performance completed on the eleventh and final day of testing. It was predicted that both fingerprint training groups would exhibit significantly greater improvement on the xQ across sessions than controls, and that the Mixed training group would display superior performance across sessions compared to the Massed training group. Instead, the results suggested that, while the Massed training group performed more accurately overall, none of the three groups improved significantly over sessions. This study has potential implications for the training of future fingerprint experts and could reduce the risk of costly errors made by these experts.

Keywords: Fingerprint expertise, forensic science, interleaved practice, perceptual learning

Word Count: 247

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.

Sept 2020

Contribution Statement

My supervisor and I collaborated to formulate relevant research questions and to design an appropriate methodology for my experiment. I completed the ethics application and pre-registered the experiment on the Open Science Framework (OSF). The fingerprint and control training tasks were designed collaboratively and, along with the xQ test of fingerprint expertise performance, were programmed by my supervisor using LiveCode Community. I ran simulations across all participants and sessions as a pilot test of the experiment. Data collection was also collaborative, with data from the Control group shared across my experiment and a similar study designed by another Honours student ('Experiment 2' on the OSF wiki page). I personally collected data from 12 of the 19 participants, with the data from the remaining seven participants collected by my supervisor and the student behind 'Experiment 2.' Pooling participants across the two experiments allowed us to run the experiment remotely and with social distancing measures in place. My supervisor guided me through conducting my analyses in R Studio and interpreting the output. I reproduced the analyses in SPSS. Financial compensation for all participants was provided by my supervisor. Finally, I completed the write-up of this thesis.

Acknowledgments

First and foremost, I would like to thank my supervisor, Dr Rachel Searston. Rachel, thank you for your unwavering passion and excitement for this project – in the face of a global pandemic, nonetheless. I am eternally grateful for the time and effort you put into programming my experiment, walking me through R Studio (which I am now a pro at!), and answering my many, many queries. I well and truly hit the supervisor jackpot!

To Jade, the mastermind behind 'Experiment 2:' I truly cannot imagine having gone through Honours without you. Thank you for listening to my constant panicking over the tiniest things (it was usually the case that you were panicking for the same reasons), and for being a beacon of positivity and encouragement. You are one in 7.8 billion (and so are your fingerprints!).

To my Social Support Wednesday girls: thank you for your social support **every day.** Our weekly laughs down in the lab gave me the boost I needed to get through each week.

To my family: thank you for supporting me unconditionally, for reading drafts, and for giving me space when I needed it. Mum, thank you for keeping the jar beside my laptop topped up with chocolate during these last few weeks – it was very much needed!

And last (but definitely not least), thank you to my family and friends who willingly participated in my experiment. You've made a valuable contribution to science!

Mixing it Up: Developing Expertise in Forensic Fingerprint Examination Using Interleaved Practice

In January 1983, Archie Williams was sentenced to life in prison for a crime he did not commit. He was exonerated by forensic fingerprint examiners after serving 36 years in prison (Innocence Project, 2019). In another landmark case, fingerprint experts linked Jerry Watson to the murder of a 61-year-old man using fingerprints found at the scene of the crime, 30 years after it had occurred (Federal Bureau of Investigation, 2012). These examples demonstrate the profound influence of expert fingerprint examination in forensic contexts. Research indicates the "miraculous" ability of these experts to accurately discriminate between fingerprints, especially when compared to novices (those without any formal training in fingerprint examination; Thompson & Tangen, 2014, p. 2; see Searston & Tangen, 2017a; Tangen, Thompson, & McCarthy, 2011; Tangen, Kent, & Searston, 2020). Few studies have, however, investigated how to most effectively train fingerprint novices to become experts. This thesis will implement one training method using a small sample of fingerprint novices and established concepts in educational psychology.

1.1 Perceptual Expertise

Contrary to common depictions of an automated process, perceptual experts such as fingerprint examiners rely on their expertise when making decisions. In a classic experiment, participants were asked to categorise a series of photographs of common objects, such as aeroplanes (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). One participant, a former aeroplane mechanic, categorised the photographs of aeroplanes on a much deeper level than most others – for example, he was able to differentiate between the engines of different types of aeroplanes. This participant was an expert on aeroplane mechanics.

Expertise has been defined by leading researchers as "consistently superior performance on a specified set of representative tasks for a domain" (Ericsson & Lehmann,

1996, p. 277), though others suggest that it cannot be explicitly defined, nor adequately measured using a single task (Dror, 2016; Gauthier, Williams, Tarr, & Tanaka 1998). Other critics reject the dichotomous nature of expertise (i.e., expert, novice) and suggest instead that expertise exists on a continuum (Shen, Mack, & Palmeri, 2014). Perceptual expertise specifically represents a superior ability to observe and categorise stimuli, as a direct result of training and experience with those stimuli (Kellman et al., 2014). As perceptual expertise develops, cognitive load decreases and categorisation becomes more automatic (Kellman et al., 2014; Tanaka, Curran, & Sheinberg, 2005).

Perceptual experts consistently perform more accurately than novices across a variety of disciplines. Expert chess players, for example, are significantly more accurate than novices at recalling positions and solving problems within games (Bilalic, McLeod, & Gobet, 2009). Facial comparison experts make significantly less errors than novices (White, Dunn, Schmid, & Kemp, 2015). Research investigating the reasons behind these superior abilities indicates that experts attend to the perceptual details that distinguish stimuli from one another, while novices tend to look at stimuli more holistically (Tanaka et al., 2005). Experts also consider the relationships between different features as well as the features themselves, known as configural processing (Gauthier et al., 1998). Facial image comparison research found that experts tend to make slower, more deliberate judgments (White et al., 2015). Finally, expertise tends not to generalise across classes of stimuli; for example, dog expertise does not generalise to bird expertise (Tanaka & Taylor, 1991; see also Bilalic et al., 2009; Gauthier, Skudlarski, Gore, & Anderson, 2000; Searston & Tangen, 2017b).

All experts were once novices. Research has demonstrated that practice with a class of stimuli can improve performance and lead to expertise (Dosher & Lu, 2005; Gauthier et al., 1998; Tanaka et al., 2005). For example, Gauthier and colleagues (1997, 1998) trained a group of novices to categorise a novel class of human-like stimuli (known as ‘Greebles’)

using a variety of categorisation, naming, and verification tasks. Following training, learners were deemed ‘experts’ and subsequently compared to a new group of novices. The experts learned novel sets of Greebles significantly faster than the novices, indicating a superior ability to transfer knowledge to new stimuli (Gauthier et al., 1998). Participants in another study were trained to categorise birds on either the basic (i.e., family) or subordinate (i.e., species) level (Tanaka et al., 2005). Those with subordinate-level training demonstrated greater transfer of learning to novel exemplars and species than those with basic-level training, providing support for the notion that experts categorise stimuli from their domain of expertise on a subordinate level (Busey & Vanderkolk, 2005; Tanaka, 2001).

While perceptual experts have demonstrated impressive levels of accuracy relative to novices, they are not immune to errors (Dror & Charlton, 2006; Searston et al., 2019). Expert decisions are influenced by previous cases, contextual details of the current case, and the quality of the perceptual information in the stimulus (Dosher & Lu, 2005; Edmond, Tangen, Searston, & Dror, 2015; Lu & Dosher, 1999). The interpretations of forensic experts were believed to be infallible until the release of two reports by the National Academy of Sciences (NAS; National Research Council, Committee on Identifying the Needs of the Forensic Science Community, 2009) and the President’s Council of Advisors on Science and Technology (PCAST; 2016). Both reports highlighted the lack of a scientific basis in many forensic disciplines and recommended empirical research into expert capabilities. The PCAST report specifically investigated pattern-matching forensic disciplines, including fingerprint examination, in which two samples are compared visually and their origin is determined by a human examiner.

1.2 Expertise in Fingerprint Examination

Fingerprint examiners have been presenting evidence in forensic investigations since the late 19th century (Faulds, 1880). These experts spend their days comparing latent prints

(those found at a crime scene) with highly similar candidate prints (those collected in a laboratory) generated by a computerised database (Dror & Mnookin, 2010). Time limits are self-imposed, with individual comparisons taking up to several hours (Busey & Vanderkolk, 2005; Shen et al., 2014). Experts arrive at one of three decisions: an individualisation (the two prints originate from the same source), an exclusion (the two prints originate from two different sources), or an inconclusive decision (the information in the latent print is insufficient to make a decision; Busey & Vanderkolk, 2005; Towler et al., 2018).

Fingerprint examination is a highly difficult task. Latent prints tend to be of low quality, and can be contaminated by dust, surface texture, and pressure (Busey & Vanderkolk, 2005; Vokey, Tangen, & Cole, 2009). Candidate prints, however, are collected in controlled environments, and typically contain more information than latent prints (Kellman et al., 2014). See Figure 1 below for a comparison between latent and candidate prints. Additionally, while all individuals have unique fingerprints, the same finger can leave varying impressions depending on pressure, perspiration, skin elasticity, and the surface the print is left on (Busey & Parada, 2010; Dror & Cole, 2010; Searston & Chin, 2019; Towler et al., 2018). The recent technological advances in fingerprint examination come at a cost; as more fingerprints are entered into the various electronic databases, the risk of making errors also increases (Cole, 2005; Dror & Mnookin, 2010). Furthermore, fatigued experts demonstrate reduced accuracy, make more inconclusive decisions, and terminate the examination process sooner (Busey, Swofford, Vanderkolk, & Emerick, 2015). Despite these caveats, expert fingerprint examiners are incredibly accurate.

Figure 1. Comparison of a latent fingerprint (left) and a candidate fingerprint (right). Sourced from Kellman et al., 2014.

1.2.1 Expert Accuracy. The NAS (2009) and PCAST (2016) reports recommended the establishment of a scientific basis for the domain of fingerprint examination. The extensive body of research that has since been conducted provides overwhelming support for the superior abilities of fingerprint experts (e.g., Searston & Tangen, 2017a; Tangen et al., 2011; Thompson & Tangen, 2014; Ulery, Hicklin, Buscaglia, & Roberts, 2011). In one of the first studies to investigate the accuracy of fingerprint experts, Tangen and colleagues (2011) presented expert and novice examiners with pairs of prints (one latent print, one candidate print) and asked them to indicate whether the prints originated from the same source (a match) or two different sources (a non-match). Experts and novices both performed well, with experts exhibiting significantly higher accuracy. Experts in another experiment outperformed novices when viewing fragments of latent prints and were more resistant to longer delays between the studied fragments and the test prints (Busey & Vanderkolk, 2005). In a recent study by Searston and Tangen (2017a), experts could accurately discriminate prints as belonging to the same or a different person, despite this task differing markedly from their everyday casework (i.e., determining whether two prints belong to the same or a

different *finger*). Experts also performed well when prints were inverted, obstructed by visual noise, or presented very briefly (Busey & Vanderkolk, 2005; Thompson & Tangen, 2014).

Though fingerprint examiners are incredibly accurate under a variety of conditions, they can make errors (Dror & Charlton, 2006; Searston et al., 2019). Applying a signal detection framework (Green & Swets, 1966), experts can make two types of errors: false positives (indicating two prints match when they do not) and false negatives (indicating two prints do not match when they do). Fingerprint examiners tend to make more false negative errors than false positive, a phenomenon known as a conservative response bias (Kellman et al., 2014; Tangen et al., 2011; Thompson, Tangen, & McCarthy, 2014; Ulery et al., 2011). Thus, fingerprint examiners are more likely to fail to identify a criminal than provide evidence that may falsely convict an innocent person (Searston, Tangen, & Eva, 2016). Given the severe potential consequences of errors made by fingerprint examiners, it is crucial to develop evidence-based training programs that result in high levels of expert performance.

1.2.2 Fingerprint Expertise Training. Despite several reports describing the need for standardised, empirically validated training programs in forensic domains, the existing literature is limited (Campbell, 2011; NAS; 2009; National Institute of Standards and Technology, 2012; PCAST, 2016). The research that does exist is of varying quality, with one study simply providing a PowerPoint presentation of direct quotes from expert examiners as their training method, before concluding that this method was insufficient in producing expertise (Stevenage & Pitfield, 2016). Forensic examiners rely on their training and experience when making decisions, so it is crucial that this training is empirically evaluated (Searston & Tangen, 2017c). Current fingerprint examiner training in Australia typically lasts around five years and involves a combination of on-the-job mentoring from qualified experts, structured lessons, and theoretical and practical examinations (Searston & Tangen, 2017d). The NAS (2009) report criticised mentor-apprentice systems, instead recommending

standardised training programs. Conducting research into the most effective ways to train fingerprint novices to become experts will result in reduced training times, less errors, and greater validity of the evidence presented in criminal courts by fingerprint experts (Mustonen & Himberg, 2011; Thompson, Tangen, & McCarthy, 2013). Even modest gains in training efficiency would be beneficial, due to the substantial amount of training required to become an expert (Roads, Mozer, & Busey, 2016).

One experiment focused on fingerprint training investigated three learning techniques with four groups of novices (Searston & Tangen, 2017c). During the training phase, a baseline group provided 'match' or 'non-match' decisions for 50 pairs of fingerprints, with a second group completing the same task while receiving feedback on their performance in each trial. The third group viewed print pairs labelled with either 'match' or 'non-match,' and listed the similarities and dissimilarities between the prints. The fourth group provided 'match' or 'non-match' decisions for two simultaneously-presented print pairs, in which a candidate print was presented twice, alongside both a matching and non-matching latent print. All three training groups demonstrated significantly greater discrimination accuracy than the baseline group, providing initial support for the implementation of the three learning techniques in fingerprint training research. However, the testing phase occurred immediately following the training phase, which could be seen as cramming. Educational research suggests that cramming can enhance initial performance but not necessarily long-term performance, as information learnt during training fades and is eventually forgotten (Tigner, 1999). Thus, the gap between the training and test phases should be maximised if the goal is to develop training procedures that result in lasting improvements in performance.

Few longitudinal studies exist in which novice performance is measured over time. One such study (Searston & Tangen, 2017d) tracked 24 trainees in a forensic laboratory over one year, testing their performance on four measures that had differentiated between expert

and novice fingerprint examiners in previous research. Significant improvement in discrimination accuracy was found across all four measures, with the majority of learning occurring during the first three months of training. This result supports the notion that learning occurs most rapidly in the early stages of training (Hawkey, Amitay, & Moore, 2004; Poggio, Fahle, & Edelman, 1992), and suggests a quadratic trend in fingerprint training performance. Future training research could expect to find a similar trend, provided an effective training method is implemented. One method that has been explored in perceptual learning research is interleaved practice.

1.3 Interleaved Practice

We classify objects into categories daily in order to structure and understand our world (Birnbaum, Kornell, Bjork, & Bjork, 2013). Inductive learning refers to the process of learning a new category by observing exemplars (Kornell & Bjork, 2008). When learning a new category, it is crucial to identify the principles, patterns, and concepts that define that category, rather than memorise specific exemplars (Birnbaum et al., 2013; Kornell & Bjork, 2008). In the context of fingerprints, examiners draw on their prior experience and knowledge of print features when making decisions (Thompson et al., 2013). Research assists us in understanding the conditions under which category learning is optimised, with several studies focusing on the presentation of exemplars.

Presenting exemplars from different categories during the learning process is referred to as interleaving, or ‘mixing’ (Carvalho & Goldstone, 2017). For example, when learning about different varieties of apples, mixing would involve presenting, say, a Red Delicious, followed by a Pink Lady, and then a Granny Smith, and so on. Alternatively, exemplars from the same category can be concentrated, or ‘massed,’ into a single block of learning (Taylor & Rohrer, 2010). With reference to the apple example, massing could involve presenting several Pink Lady apples in one learning block, with the next block featuring several Granny

Smith apples, and so on. While early interleaving research investigated the learning of motor skills (e.g., Hall, Domingues, & Cavazos, 1994), research has since been conducted using the styles of artists' paintings (Kang & Pashler, 2012; Kornell & Bjork, 2008; Zulkiply & Burt, 2013), bird species (Birnbaum et al., 2013; Wahlheim, Dunlosky, & Jacoby, 2011), mathematical problems (Mielicki, 2019; Rohrer, 2012; Taylor & Rohrer, 2010) and simulated blob figures (Carvalho & Goldstone, 2014a, 2014b).

Research has demonstrated a benefit of mixing exemplars in perceptual learning tasks. In one experiment, Kornell and Bjork (2008) tested participants' ability to classify paintings by a group of 12 artists. During the learning phase, the paintings were either mixed (i.e., a painting by one artist was followed by a painting by a different artist) or massed (i.e., a painting by one artist was followed by another painting by that same artist). The classification test involved previously unseen paintings by the same 12 artists and resulted in significantly greater accuracy following mixing (compared to massing). Benefits of mixing exemplars have also been found among participants learning to classify species of butterflies (Birnbaum et al., 2013) and simulated blob figures (Carvalho & Goldstone, 2014a, 2014b).

The Discriminative-Contrast Hypothesis (DCH; Goldstone, 1996) purports that mixing benefits learning by juxtaposing exemplars from different categories, highlighting the salient differences *between* those categories. An attentional focus on between-category differences is likely to result in better encoding and memory for such differences and, in turn, enhanced classification accuracy (Carvalho & Goldstone, 2017). Bird expert David Sibley (2002) explains the practical importance of identifying between-category differences: "Whether you are looking at two birds side by side in the field or comparing a bird in the field to pictures in a book, you must make comparisons and search for differences" (p. 22). The DCH also purports that massing encourages the identification of similarities *within* categories (Birnbaum et al., 2013; Carvalho & Goldstone, 2014a; Goldstone, 1996; Kornell &

Bjork, 2008). In support of the DCH, Carvalho and Goldstone (2017) found that, following massed study, learners recalled characteristic features (within-category similarities) of studied exemplars more accurately than discriminative features (between-category differences). Furthermore, eye tracking studies confirmed that learners paid more attention to discriminative features during mixed study, but not during massed study (Carvalho & Goldstone, 2017).

1.3.1 The Optimal Conditions for Interleaved Practice. Both mixing and massing are potentially beneficial for inductive learning under different conditions. Mixing has demonstrated effectiveness with low-discriminability categories, or those with many between-category differences (Carvalho & Goldstone, 2014a, 2014b; Kang & Pashler, 2011; Kornell & Bjork, 2008; Kurtz & Hovland, 1956; Zulkiply & Burt, 2013). Massing, on the other hand, is more likely to be effective with high-discriminability categories, or those in which exemplars share many within-category similarities (Carvalho & Goldstone, 2014a; Goldstone, 1996; Kurtz & Hovland, 1956; Zulkiply & Burt, 2013). For example, Zulkiply and Burt (2013) presented learners with two sets of stimuli that were designed to be of either high- or low-discriminability. An advantage of mixing for low-discriminability categories was found, along with an advantage of massing for high-discriminability categories. The key finding in such research is that the rapid alternation of categories when mixing allows the identification of between-category differences, which is particularly beneficial for learning when such differences are hard to detect (Carvalho & Goldstone, 2014a).

In addition to enhancing learning with low-discriminability categories, research has demonstrated a benefit of presenting mixed exemplars simultaneously (i.e., multiple exemplars appearing at once) as opposed to sequentially (i.e., one exemplar at a time). Past experiments have presented exemplars (e.g., images of birds, paintings) simultaneously under both mixed and massed conditions, finding a benefit for mixed conditions only (compared to

sequential presentation; Kang & Pashler, 2012; Wahlheim et al., 2011). Perhaps most importantly, Carvalho and Goldstone (2014a) demonstrated that simultaneous-mixed presentation (i.e., multiple exemplars from different categories presented at once) resulted in a performance advantage for low-discriminability categories. This is a pertinent finding in relation to fingerprint expertise training, as fingerprints are difficult to distinguish from one another, especially for novices (e.g., Tangen et al., 2011; Thompson & Tangen, 2014). Incorporating a simultaneous-mixed presentation schedule, therefore, may have some benefit with fingerprints.

Several studies have demonstrated a tendency of mixing to impair initial performance, but lead to superior retention on delayed tests (Lin et al., 2011; Mayfield & Chase, 2002; Rohrer, 2012; Rohrer & Taylor, 2007; Taylor & Rohrer, 2010). For example, when types of mathematical problems were mixed, both primary school children (Taylor & Rohrer, 2010) and university students (Rohrer & Taylor, 2007) demonstrated impaired performance during a practice phase and enhanced accuracy on a later test, compared to when the problems were massed. This delayed retention benefit may be due to the retrieval of knowledge or exemplars from long-term memory stores required by mixing; the greater cognitive effort required makes mixing a "desirable difficulty" (Bjork, as cited in Kornell & Bjork, 2008, p. 586). The delayed retention benefit also demonstrates the distinction between learning and performance, in that interleaving may impair performance and benefit learning (Lin et al., 2011; Soderstrom & Bjork, 2015).

1.4 The Current Study

While there exists a wealth of literature on both fingerprint examination and interleaved practice, only one study has, thus far, combined the two areas of research. During each trial of a brief training phase, Searston and Tangen (2017c) presented novices with two images of the same candidate print, mixed in with both a matching and a non-matching latent

print. This resulted in significantly greater classification accuracy on a later test than a baseline group viewing just one fingerprint at a time. This study did not include a pre-test of fingerprint examination ability, which would have brought to light any initial individual differences, nor an active control group, which would have allowed the assessment of the effect of fingerprint training in general, compared to no fingerprint training. These are research gaps I intend to fill in this thesis. Research suggests that mixing has optimal benefit when exemplars from low-discriminability categories (e.g., fingerprints) are presented simultaneously, and participants are tested on their classification performance following a delay. The current experiment was designed in a way that reflects these conditions. Research also suggests that fingerprint training may result in a quadratic trend in performance over time, in that learning rapidly increases in the initial stages of training and gradually plateaus (Searston & Tangen, 2017d).

In this thesis, I implement a novel training protocol with three groups of fingerprint novices. Performance on a test of fingerprint expertise will be recorded over 11 sessions on 11 consecutive days. Participants will also complete 10 training sessions as per their allocated condition. Two fingerprint training groups (Mixed, Massed) will compare sets of four simultaneously-presented prints (three latent prints and one candidate print), and decide whether one randomly highlighted latent print matches, or does not match, the candidate print. The three latent prints in a trial of the Mixed training task will all be sourced from different fingers, while those in a trial of the Massed training task will be sourced from the same finger. Control participants will complete a task unrelated to fingerprints. Based on the extensive literature on interleaved practice with complex categories, I predict the following:

1. Participants in both the Mixed and Massed training groups will demonstrate a significant quadratic learning benefit over sessions, compared to the Control training group.

2. Participants in the Mixed training group will demonstrate a significantly greater quadratic learning benefit over sessions, compared to the Massed training group.

CHAPTER 2 - METHOD

My experiment was preregistered on the Open Science Framework (OSF) under 'Experiment 1' and can be found here. The wiki page includes a detailed description of the research aims, predictions, methodology, and analysis plan, with links to the R markdown scripts. The page also describes a similar experiment conducted by another Honours student within the School of Psychology at the University of Adelaide (Experiment 2). Pre-registering experiments assists in fostering a culture of openness within the research community, allows the verification and reproducibility of findings, and lessens the impact of hindsight bias (Searston et al., 2019).

2.1 Ethics Statement

Ethics approval was granted by the University of Adelaide's Human Research Ethics Sub-Committee (HREC 20/31).

2.2 Experimental Parameters

Novice participants were assigned to one of three conditions, with two groups undertaking training in fingerprint examination and the third serving as an active control group.

2.2.1 Fingerprint Training Protocol. Each fingerprint training session involved 100 trials, with each trial consisting of three latent prints presented alongside one candidate print. One of the latent prints was randomly highlighted with a bright purple outline, with participants asked to consider all four prints and decide whether the highlighted print and the candidate print (which was also highlighted) originated from the same finger, or two different fingers. There was no time limit for individual trials, however a warning message appeared whenever a participant took longer than 10 seconds to respond. Participants received feedback on every trial, with an audible low-key tone and red highlights in response to incorrect decisions, and a high-key tone and green highlights following correct decisions.

Two groups (Mixed, Massed) completed this task on each of 10 consecutive days, with the only difference between the groups lying in the prints they responded to. The three latent prints in each trial of the Mixed training task originated from three different fingers, with the highlighted print 'mixed' in. Each trial in the Massed training task involved a latent print 'massed' with two other prints from the same finger. As per interleaving theory, the Mixed training task displayed stimuli from different categories (i.e., different fingers) and the Massed training task displayed stimuli from the same category (i.e., the same finger). Altogether, the 100 trials took approximately 15 minutes to complete regardless of condition. The task was kept as brief as possible to reduce the potential effects of fatigue on performance, while still providing a reasonable amount of training (Busey et al., 2015). Having both groups complete the same task with different stimuli allowed me to make inferences between the type of training and performance across sessions. See Appendix A for an example of a trial of the fingerprint training tasks.

2.2.2 Control Training Protocol. Participants in the Control training group completed a word jumble task, with each trial consisting of one jumbled word (the target) presented alongside the unjumbled target word and four highly similar distractor words. The task was to identify the unjumbled target word as quickly as possible. All words were sourced from an online repository of 1,371 common seven-letter words (GitHub, 2015). Like the fingerprint training tasks, this task consisted of 100 trials and took approximately 15 minutes to complete. A warning message appeared when participants took longer than 10 seconds to respond. Participants received feedback on their performance in each trial. The intention was to have the control group complete a task that was of a similar duration to the training interventions and not perceptually based, such that I could make inferences regarding the effect of the fingerprint training tasks, and perceptual training more broadly. Word jumble tasks have previously been used as filler tasks and measures of performance (e.g., Dunphy &

Milbourne, 2009; Rinck & Becker, 2005; Tran, Siemer, & Joormann, 2011). See Appendix B for an example of a trial of the Control training task.

2.3 Measures

2.3.1 The xQ. The Expertise Quotient or "xQ" (Searston, Tangen, & Thompson, in preparation) for fingerprints consists of three tasks: Print Matching, Print Recognition, and Print Nomination. The tasks require participants to determine whether two prints originate from the same or different sources (Print Matching), hold a briefly-presented test print in their working memory and sort through a series of prints to select the matching one (Print Recognition), and determine which hand (i.e., left or right) and finger (i.e., thumb, index etc) a print originates from (Print Nomination). The three tasks were part of an initial set of ten tasks that captured multiple and varied components of the fingerprint examination process and included 603 cases that took over 3 hours for fingerprint examiners to complete. During the task refinement process, all ten tasks were completed by 44 Australian fingerprint experts and 44 student novices. All possible combinations of the ten tasks were computed, and the three that most optimally discriminated between experts and novices were identified. Further refinement involved computing all possible combinations within the top three tasks to identify the number of trials needed to optimally classify experts and novices. The final xQ takes approximately 12 minutes to complete and comprises 34 trials: 12 Print Matching trials, 12 Print Recognition trials, and 10 Print Nomination trials. No feedback is provided on any of the trials in any of the tasks. Scores on the xQ are calculated by averaging participants' scores on each of the three tasks and are presented as a proportion of correct responses to total responses (i.e., a score of .43 indicates that 43% of responses were correct). The software also records participant response times and self-reported confidence levels (i.e., how many trials they believe they answered correctly on each task). This thesis is one of the first studies to

implement the xQ as a measure of fingerprint expertise performance and could provide validatory support for its use in future research.

2.4 Materials

All participants completed the experiment on a 13-inch Apple MacBook Pro.

2.4.1 Fingerprint Images. The fingerprint images used in the training tasks consisted of sets of nine prints from the right thumbs of 25 individuals, with each set containing five latent and four candidate prints. Two of the latent prints in each set were left on a metal surface, two on a glass surface, and one on a timber surface. Each print was 600 millimetres by 600 millimetres but was presented at 352 by 352 millimetres such that all four prints fit comfortably on the screen. All of the original details in the prints were left intact (e.g., natural variation in contrast, hue, luminance).

Each of the 100 candidate prints appeared in the 100 trials of the fingerprint training tasks (4 candidate prints for 25 individual fingers) but were randomly sampled in a different order for each session, for each participant. They were also presented alongside different latent prints in each trial, such that the correct response (as indicated by the auditory and visual feedback) in one session was unrelated to the correct response in any other session. Latent prints were presented on the left side of the screen, with the candidate print on the right, reflecting the layout of examiner casework (Roads et al., 2016). In the Massed training task, we randomly sampled three out of the five latent prints from one of the 25 individual fingers for each trial, and for each participant. In this condition, the latent prints were sampled from the same finger as the candidate print on a random half of the trials and from a randomly sampled finger on the remaining trials (i.e., half matching trials, half non-matching trials). In the Mixed training task, we randomly sampled one out of five latent prints from three of the 25 individual fingers for each trial and participant. None of the fingerprint images

used in the training tasks were used in the xQ; thus, any improvement observed over the 11 sessions cannot be attributed to learning the specific training images.

The fingerprint images in the xQ were taken from a large set of prints that included approximately 10 unique prints from 300 individuals. In each session, for each participant, fingerprints were randomly selected (with replacement). There were a few repeated prints in the xQ across the 11 sessions, though performance cannot be attributed to particular prints due to their random selection.

None of the fingerprint images are available on the OSF due to ethical and legal constraints put in place to protect the identities of the sources of the fingerprints (Searston et al., 2019). This was also the reason for the experiment being completed in person, rather than through an online survey link.

2.5 Software

The xQ test of fingerprint expertise, fingerprint training tasks, and word scramble tasks were all programmed by my supervisor using LiveCode Community (version 9.5.0). Prior to data collection, I conducted simulations for all participants and sessions as a pilot test of the experiment. I inspected the simulated data files using Apple's Numbers (version 10) application. The data were analysed using both R Studio (version 1.3.1056) and SPSS (version 27).

2.6 Participants

As a result of this study being conducted during a global pandemic with social distancing restrictions in place, participants formed a convenience sample of family, friends, and network connections of my supervisor and I (N = 19). The sample included 13 females and 6 males, ranging from 19-62 years old (M = 32.4). Participants were randomly assigned to one of three conditions, with six participants in each fingerprint training condition and seven in the Control condition. The gender split was approximately equal across the three

conditions, however the mean ages of the groups differed slightly, with an average of 25 years in the Control group, 42 years in the Mixed training group and 31.5 years in the Massed training group.

All participants were required to be at least 18 years of age, be fluent in English, have normal or corrected-to-normal vision, and have no formal experience in fingerprint examination. Recruitment was on a voluntary basis and occurred in person or online. Compensation was provided in the form of a $20 gift card per hour of participation, with $80 provided upon completion of the full experiment. Responses were to be excluded if participants failed to complete the experiment, took less than 500 milliseconds to respond to more than 30% of trials, or provided the same response to more than 80% of trials. None of the participants met these criteria.

Given the limited participant pool available due to the global pandemic, I was a participant in my own experiment and was randomly assigned to the Massed training condition. Ideally, all participants would be completely blind to the true aims and hypotheses of experiments (Rosenthal, 1966). I conducted the full analyses both with and without my own data, with the pattern of results remaining unchanged.

We had aimed to collect data from 18 participants (six in each of the three conditions), but, due to a coding error, collected data from one extra participant, who was randomly assigned to the Control condition. I conducted the analyses with and without the extra participant and found no change in the pattern of results. See Appendix C for the analyses excluding my own data, and that of the extra participant, respectively.

2.6.1 Shared Controls. This experiment and Experiment 2 (as described on the OSF wiki page) implemented a shared control group, due to sampling constraints resulting from the global pandemic, and the similarity of the two experiments. Thus, the task completed and

data collected from participants in the Control training group were identical across both experiments.

2.6.2 Power Analysis. No previous research has investigated the effect of an interleaved training protocol on fingerprint expertise performance using a pre-post design and an active control group. Studies implementing interleaved practice with visual categories have produced effect sizes ranging between .70 and 1.41 (e.g., Birnbaum et al., 2013; Carvalho & Goldstone, 2014a; Kang & Pashler, 2012; Kornell & Bjork, 2008). Thus, I determined the target sample size of the current study using the Smallest Effect Size of Interest (Lakens, Scheel, & Isager, 2018) of .70. With a sample of 18 participants[1] and 374 observations per participant (34 trials in each of 11 sessions of the xQ), the experiment had an estimated power of 82.2% to detect an effect of the training conditions on xQ performance across sessions. Jake Westfall's PANGEA software (Westfall, 2020) was used to calculate the statistical power of the experiment.

2.7 Design

This experiment implemented a 3 (Condition: Mixed, Massed, Control) by 11 (Session: Session 1, Session 2, Session 3… Session 11) mixed factorial design. Condition served as a between-groups independent variable, with participants completing one condition each. All participants were assessed at all 11 time points, making Session a within-groups independent variable. Performance on the xQ served as the dependent variable.

2.8 Procedure

Participants were briefed about their involvement via the participant information sheet (see Appendix D), signed a consent form (see Appendix E), and provided basic demographic data (gender and age). Each participant was allocated a unique pseudonym known only to my supervisor and I. Initial instructions stated that participants were required to complete four

[1] Note that the 19th participant was not planned, and thus was not included in the power analysis.

tasks, taking approximately 20 to 30 minutes to complete altogether. Informing participants they were completing four tasks (rather than a test, and a training task) meant that they allocated attention to each of the four tasks equally (Roads et al., 2016). Task-specific instructions appeared immediately prior to each task. Screenshots of the instructions provided to participants can be found in Appendix F. All participants completed 11 sessions across 11 consecutive days, with the first 10 sessions consisting of the xQ (tasks one to three) followed by their assigned training task (task four). The eleventh session involved one final attempt of the xQ, but not the training task. The sessions were completed at the same time each day, ensuring there was a 24-hour gap between each training session and the subsequent xQ attempt; this minimised the cramming effect (Tigner, 1999). Furthermore, the delay allowed us to investigate the notion that mixing exemplars impairs initial performance, but enhances long-term retention (e.g., Rohrer, 2012; Rohrer & Taylor, 2007; Taylor & Rohrer, 2010). Data collection ceased when all 19 participants had completed the experiment. Following completion, participants received their payment and were debriefed regarding the true nature of the experiment.

CHAPTER 3 – RESULTS

Overall, participants appeared to perform relatively poorly on the xQ, with minimal improvement occurring between sessions one (M = .420, SD = .112) and 11 (M = .441, SD = .089). The Massed training group (M = .451, SD = .082) appeared to achieve greater performance on the xQ, collapsed across sessions, than the Mixed training group (M = 0.426, SD = .089), with the Control training group (M = .416, SD = .084) demonstrating the lowest performance overall. See Table 1 below for the means and standard deviations of performance on the xQ, for each condition and in each session.

Table 1.
Descriptive statistics for performance on the xQ.

	Mean Performance on the xQ		
Session	Control (SD)	Mixed (SD)	Massed (SD)
1	.392 (.118)	.434 (.099)	.438 (.128)
2	.378 (.081)	.387 (.054)	.409 (.078)
3	.401 (.074)	.398 (.120)	.442 (.071)
4	.398 (.068)	.412 (.067)	.489 (.066)
5	.390 (.028)	.388 (.116)	.434 (.060)
6	.439 (.139)	.425 (.103)	.431 (.077)
7	.452 (.029)	.425 (.105)	.476 (.036)
8	.417 (.091)	.444 (.060)	.490 (.107)
9	.471 (.098)	.426 (.093)	.455 (.095)
10	.432 (.066)	.479 (.068)	.445 (.082)
11	.410 (.080)	.466 (.097)	.453 (.095)

Performance on the xQ for each of the three training groups across all 11 sessions is represented visually in Figure 2 below. The graph suggests that there was very little, if any, improvement on the xQ across sessions for all three training groups. It also suggests that there was a moderate level of variability in performance between participants, across sessions.

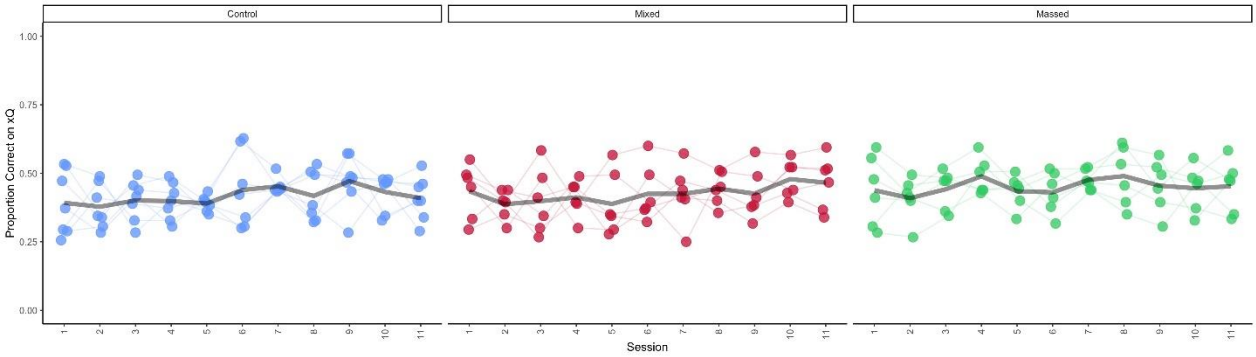


Figure 2. Performance on the xQ across sessions for the Control (blue), Mixed (pink) and Massed (green) training groups. Black lines represent the mean performance of each of the three groups across sessions. Each individual dot represents the score of one participant on one session of the xQ.

3.1 Confirmatory Analysis

As planned in my pre-registration, I conducted a 3 (Condition: Mixed training, Massed training, Control training) by 11 (Session: Session 1, Session 2, … Session 11) mixed factorial ANOVA to investigate the performance of a sample of fingerprint novices on the xQ, over sessions. The xQ scores were found to be normally distributed, with skewness of .012 and kurtosis of -.654. See Appendix G for a histogram of the xQ data. The mean variances over sessions were approximately equal between the Control (σ^2 = .007), Mixed (σ^2 = .008) and Massed (σ^2 = .007) training conditions. The data in Figure 2 provide further evidence that the variances between groups, across sessions, were similar. An outlier was identified in session seven of the xQ, but I found this to represent a genuine variation in performance and proceeded with the analysis. All statistics reported below are two-tailed. Effect sizes are reported using Cohen's (1988) conventions of small (>.1), medium (>.3), and large (>.5) effects. All effects are reports as significant at $p < .05$, unless otherwise stated.

The ANOVA revealed non-significant main effects of both Condition ($F(2, 16) = .879$, $p = .434$, $\eta^2_G = .032$) and Session ($F(10, 160) = 1.463$, $p = .158$, $\eta^2_G = .060$). The interaction effect of Condition and Session was also non-significant ($F(20, 160) = .549$, $p = .941$, $\eta^2_G = .046$). Mauchly's W was non-significant, indicating that the assumption of

sphericity was met (W = .014, p = .570). See Appendix H for the full output of the ANOVA as conducted in R Studio.

3.1.1 Test of Hypothesis 1. I predicted that both the Mixed and Massed training groups would demonstrate a significant improvement in performance on the xQ over sessions, compared to the Control training group. Treatment-control contrasts revealed that the Massed training group achieved significantly greater accuracy than the Control training group (t = 2.367, p = .019), but the Mixed training group did not (t = .641, p = .522). A subsequent trend analysis revealed no significant quadratic trend over sessions (t = -.885, p = .378).

3.1.2 Test of Hypothesis 2. I also predicted that the Mixed training group would demonstrate significant improvement on the xQ over sessions compared to the Massed training group. Helmert contrasts demonstrated that the Massed training group achieved significantly greater overall performance than both the Mixed and Control training groups (t = 2.305, p = .022). A significant linear trend was also found (t = 2.573, p = .011). However, a similar trend was not found in the treatment-control contrasts, reducing the robustness of this finding.

3.2 Exploratory Analysis

After analysing the xQ data, I was interested in whether any, or all, of my training groups demonstrated improvement on the training tasks across sessions. The Control training group completed an entirely different task to the two fingerprint training groups and was excluded from the following analysis.

Participants in the Massed training group (M = .668, SD = .088) appeared to perform more accurately on their training task compared to the Mixed training group (M = .582, SD = .058), collapsed across sessions. The Massed training group also seemed to demonstrate more

improvement between sessions one (M = .577, SD = .034) and ten[2] (M = .703, SD = .112) than the Mixed training group (session one: M = .593, SD = .045, session ten: M = .563, SD = .063). See Table 2 below for the means and standard deviations of performance on the fingerprint training tasks across sessions.

Table 2.
Descriptive statistics for performance on the two fingerprint training tasks.

	Mean Performance on the Fingerprint Training Tasks	
Session	Mixed (SD)	Massed (SD)
1	.593 (.045)	.577 (.034)
2	.572 (.040)	.648 (.112)
3	.580 (.094)	.665 (.055)
4	.592 (.062)	.647 (.074)
5	.585 (.050)	.652 (.093)
6	.587 (.023)	.695 (.067)
7	.578 (.083)	.703 (.084)
8	.600 (.070)	.668 (.101)
9	.568 (.060)	.722 (.089)
10	.563 (.063)	.703 (.112)

Performance across sessions on the Mixed and Massed training tasks is displayed in Figure 3 below. The graph suggests no improvement on the Mixed training task, and a small improvement on the Massed training task across the 10 training sessions.

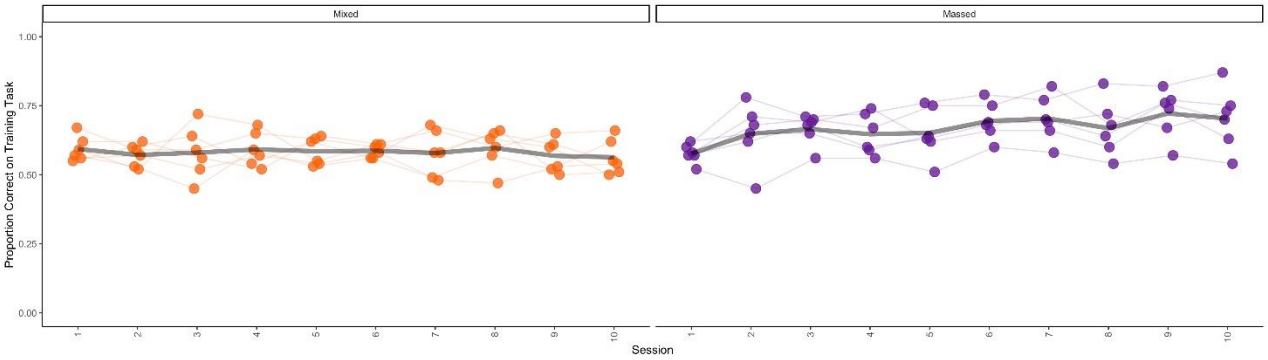


Figure 3. Performance of the Mixed (orange) and Massed (purple) fingerprint training groups on the training tasks across sessions. Black lines represent the mean scores of each training group across sessions. Each individual dot represents the score of one participant in one training session.

[2] Note that, while participants completed 11 sessions in total, the training tasks were only completed during the first ten sessions.

A mixed factorial ANOVA revealed a small but significant main effect of Condition ($F(1, 10) = 6.664$, $p = .027$, $\eta^2_G = .286$) and a non-significant main effect of Session ($F(9, 90) = 1.585$, $p = .132$, $\eta^2_G = .060$). The interaction between Condition and Session was significant ($F(9, 90) = 2.920$, $p = .004$, $\eta^2_G = .104$), though it yielded a small effect. Mauchly's W was non-significant ($W = .000$, $p = .301$), indicating that the assumption of sphericity was met. Treatment-control contrasts suggested that the Massed training group performed significantly more accurately on their training task than the Mixed training group, across sessions ($t = 6.336$, $p < .001$). A subsequent trend analysis found a significant linear trend over sessions for the Massed training group, compared to the Mixed training group ($t = 2.718$, $p = .008$). See Appendix I for the full output of the ANOVA as conducted in R Studio.

CHAPTER 4 – DISCUSSION

This thesis aimed to investigate the performance of three training groups (Mixed, Massed, Controls) on a test of fingerprint expertise, the xQ, over multiple sessions. I predicted that the two fingerprint training groups (Mixed, Massed) would demonstrate a significant learning benefit across sessions compared to controls, and that the Mixed training group would significantly outperform the Massed training group over sessions. Neither hypothesis was fully supported, with the results indicating superior overall performance on the xQ for the Massed training group, and no significant improvement over sessions for any of the three groups. An exploratory analysis indicated that the Massed training group significantly improved over sessions on the training task. Potential explanations for these findings are outlined below.

4.1 Confirmatory Analysis

Contrary to my predictions, the Massed training group demonstrated superior performance on the xQ compared to both the Mixed and Control training groups. Despite this superior performance, the Massed training group did not significantly improve on the xQ over sessions. These results may be partly due to chance. It is possible that the most accurate performers were randomly assigned to the Massed training group during the allocation process. Previous research suggests that initial novice performance in fingerprint examination is highly variable, and that individual differences in accuracy tend to persist, especially when task demands vary (Searston & Tangen, 2017d). Fingerprint examination is thought to be a variable task due to the novelty of the prints in each case (Ackerman, 1987). Initial individual differences in performance on the xQ seemed to persist in this experiment, in that the Massed training group consistently performed more accurately than the other two training groups. Future fingerprint training research could provide measures of initial individual differences to

determine whether these are approximately equal across training groups; for example, baseline measures of basic pattern-matching skills, and other cognitive abilities.

While the Massed training group demonstrated consistently superior performance on the xQ compared to the Mixed and Control training groups, they did not significantly improve over sessions. They did, however, improve significantly on the training task, demonstrating the specificity of expertise. Learning on a trained task does not necessarily generalise to learning on an untrained task (in this case, the xQ), even when the same stimulus is used in both tasks (Sagi, 2011; Shiu & Pashler, 1992; Wright & Zhang, 2009). For example, trainees successfully learnt to discriminate the brightness of lines, but this training did not improve their ability to discriminate the orientation of the same lines (Shiu & Pashler, 1992). It is suggested that specificity of learning is more common with demanding, high-precision tasks (i.e., those that require discrimination between highly similar stimuli; Ahissar & Hochstein, 1997; Jeter, Dosher, Lu, & Petrov, 2009; Liu & Weinshall, 2000). Fingerprints are highly similar to one another and difficult to discriminate between, which may explain why the improvement on the training task did not transfer to improvement on the xQ.

I predicted that the two fingerprint training groups would demonstrate a significant quadratic trend, or traditional learning curve, over sessions, in that performance on the xQ would rapidly increase over the first few sessions and gradually plateau. However, no significant trends were found. This prediction was based on research by Searston and Tangen (2017d) in which 24 fingerprint trainees were tested over a period of one year, and demonstrated the majority of their learning in the first three months. This boost in learning may have been partly due to the trainees' increased motivation to perform well, given that their training was a mandatory part of their employment. Additionally, these trainees were working full-time, while the novices in my experiment received only three hours of training

spread across 10 days. This may have been too brief of a training period to produce a learning curve such as that demonstrated by Searston and Tangen (2017d).

Contrary to my predictions, none of the three training groups demonstrated significant improvement on the xQ over sessions. Participants received approximately three hours of training across ten sessions, which may have been insufficient in producing a learning benefit with fingerprints, a complex, low-discriminability category of stimuli. Novices in Gauthier and Tarr's (1997) research reached a pre-specified expertise criterion with a novel set of human-like stimuli known as Greebles after seven to 10 hours of training. While we were seeking to establish perceptual learning rather than perceptual expertise in this experiment, the amount of training we provided was evidently insufficient to produce a boost in performance on the xQ.

It seems likely that three hours of training on one task was insufficient in inducing a learning benefit on the xQ. Roads and colleagues (2016) noted the time-consuming nature of developing expertise in complex visual tasks such as fingerprint examination. In a recent study by Thompson and colleagues (2014), novices performed just as accurately on a fingerprint discrimination task as trainees with up to six months of training and experience. Considering my results and the lack of longitudinal research in the domain of fingerprint expertise development, future research could extend the training period to several weeks or months. However, this is a costly alternative and may incur higher attrition rates due to the long-term commitment. To mitigate this effect, researchers could recruit enrolled fingerprint trainees to ensure participants are motivated and committed (Searston and Tangen 2017d).

Extending the length of the training period should occur concurrently with the inclusion of multiple training tasks. While using just one task allowed me to assess the effect of the task, it was likely insufficient to produce learning on the xQ on its own. Research has demonstrated a benefit of interleaving task type, rather than individual exemplars (Rau,

Aleven, & Rummel, 2013; Szpiro, Wright, & Carrasco, 2014). For example, a learning benefit was found when two perceptual tasks that did not influence learning on their own were mixed (Szpiro et al., 2014). In order to induce a significant boost in learning, multiple and varied fingerprint training tasks should be interleaved in future training studies that last several weeks or months.

4.2 Exploratory Analysis

Statistical analysis of performance on the fingerprint training tasks (excluding the Control training group, who completed an unrelated task) revealed that the Massed training group performed significantly more accurately than the Mixed training group, and improved significantly across sessions. While I did not form any predictions regarding performance on the training tasks, much of the literature on interleaved practice would suggest that the Mixed training group should have demonstrated superior performance and improvement (e.g., Carvalho & Goldstone, 2014a, 2014b; Kang & Pashler, 2011; Kornell & Bjork, 2008; Zulkiply & Burt, 2013). Potential explanations for my findings are explored below.

The two fingerprint training tasks both involved 100 trials in each session, with the highlighted latent print matching the candidate print in a random half of the trials, and not matching in the other random half of trials. Consequently, the participants in the Massed training condition viewed four prints from the same finger on half of the training trials. This may have incidentally increased the amount of information available across the four prints and provided participants with more opportunities to notice similarities within prints. Noticing similarities within categories has previously been associated with learning benefit with some stimuli, potentially explaining the improvement demonstrated by the Massed training group (Carvalho & Goldstone, 2014a; Carvalho & Goldstone, 2017; Goldstone, 1996; Zulkiply & Burt, 2013). However, this was not expected with fingerprints, a complex, low-discriminability set of stimuli that should benefit from mixing exemplars (Carvalho &

Goldstone, 2014a; Goldstone, 1996). This explanation also assumes that participants noticed that the prints were all from the same source.

In the non-matching trials of the Massed training task, the prints were somewhat interleaved in that the three matching latent prints were 'mixed' in with one non-matching candidate print. Thus, this training task could be classified as a hybrid training schedule that involved both mixing (non-matching trials) and massing (matching trials). Previous research has indicated a learning benefit of hybrid schedules, especially with complex tasks (Carpenter & Mueller, 2013; de Croock & van Merrienboer, 2007; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Rau et al., 2013; Rohrer, 2012). Several studies have indicated that mixing stimuli (e.g., word lists, mathematical problems) was only beneficial for learning following a period of massing the same stimuli (Rau et al., 2014; Sorensen & Woltz, 2016). It is possible that fingerprint examination is a highly complex task that requires a hybrid training schedule incorporating both mixing and massing. The fact that the Massed training task incidentally served as a hybrid schedule may have contributed to the superior improvement demonstrated in my experiment. The effectiveness of hybrid training schedules with fingerprint examination should be investigated in future research.

The two fingerprint training tasks required participants to consider all four prints in each trial and determine whether or not the randomly highlighted latent print matched the candidate print. However, it is plausible that participants may not have considered all four prints, instead just comparing the two highlighted images. This would have made the distinction between the Mixed and Massed tasks somewhat redundant. Melton (1970) found that mixing had no effect when participants failed to recognise that exemplars were being mixed. A future study could replicate my experiment and incorporate eye tracking technology to determine whether participants view all four prints before responding. Such technology has

been previously implemented in both fingerprint training (Roads et al., 2016) and interleaved practice research (Carvalho & Goldstone, 2017).

4.3 The xQ as an Assessment Tool

The novice participants in my experiment performed relatively poorly on the xQ overall, with the three training groups consistently achieving scores of less than 50%. These results were expected, given that fingerprint examination is a highly difficult task (Busey & Vanderkolk, 2005), especially for novices (e.g., Searston & Tangen, 2017a; Tangen et al., 2011; Thompson & Tangen, 2014). The results also provide validatory support for the xQ, a recently-developed measure of fingerprint expertise performance (Searston, Tangen, & Thompson, in preparation). The xQ is intended to discriminate between fingerprint experts and novices, and the poor performance of my novice participants suggests that it does so reasonably well. Additionally, the lack of improvement over sessions demonstrated by the Control training group suggests that the xQ is, on its own, an ineffective method of inducing learning with fingerprint novices. This finding provides further evidence of the need for empirically validated training programs. In addition to assessing the performance of novices over time in fingerprint training studies, the xQ could potentially be used during the recruitment of fingerprint trainees in forensic laboratories, or to assess the abilities of current practicing examiners to ensure they are performing at the appropriate standard.

4.4 Strengths

Studies involving fingerprint novices tend to recruit undergraduate university students, who are usually of a similar age and educational background (e.g., Roads et al., 2016; Searston & Tangen, 2017c; Tangen et al., 2011). My thesis, due in part to the constraints placed on sampling during a global pandemic, employed a much broader sample, with the ages of my participants ranging between 19 and 62 years (M = 32.4). This wider age range allows me to generalise my results to a larger population, compared to if I had recruited

solely university students. The caveat with having a broad age range is that there may be age differences in fingerprint expertise performance. While I did not assess this here, future research could investigate age differences in the abilities of fingerprint novices, and whether fingerprint examination ability decreases in older age like many other cognitive functions (Murman, 2015).

Many studies implementing interleaved practice limit their learning phase to a single session, and have only a brief delay (i.e., less than one hour) between the learning and test phases (Rohrer, 2012). A major strength of my experiment was that there were 11 testing sessions, and a 24-hour gap between each training session and the subsequent xQ attempt. This assisted in reducing the effects of cramming (Tigner, 1999), which, in turn, allowed me to assess whether any genuine learning occurred.

4.5 Limitations

In addition to the limitations that have already been discussed (i.e., the insufficient length and variety of the training program, the unintended consequences of the training task design), my experiment had low fidelity. Fidelity refers to "the degree of similarity between experimental conditions and the reference domain" (Searston et al., 2016, p. 51). My training tasks required participants to compare four prints and provide a response within 10 seconds. In reality, examiners spend up to several hours comparing just two prints. Furthermore, the base rates of matching and non-matching prints in my training task (i.e., half matching trials and half non-matching trials) do not reflect the base rates of examiner casework (Thompson et al., 2013). The fidelity of my experiment was necessarily reduced by the high level of control I exerted. I specifically manipulated certain variables (e.g., the number of sessions, the various training conditions) in order to empirically assess the impact of interleaved practice on fingerprint expertise performance. The key challenge for researchers is to find the

right balance of fidelity, generalisability and control needed to answer the research question (Searston et al., 2016; Thompson et al., 2013).

Furthermore, despite calculating that my experiment had sufficient power to detect an effect of the training interventions on xQ performance, my sample size was relatively small. Future studies should aim to recruit a larger sample, which will reduce the effects of the variability in performance demonstrated by novices, and the influence of outliers (e.g., one-off performances in particular sessions; Eva, Rosenfeld, Reiter, & Norman, 2004).

4.6 Implications

Expert fingerprint examiners spend their days making high-stakes decisions in criminal investigations. While the superior accuracy of these experts is evident in the literature, errors may still occur. These errors have the potential to result in the conviction of an innocent person, or to allow a genuine perpetrator to roam free. Such serious consequences necessitate further research into the decision-making processes of fingerprint experts, and the most effective methods of training novices to become experts. Implementing training methods that have been rigorously evaluated in controlled research environments will reduce training times and the risk of making errors. While the current study did not explicitly demonstrate improvement in fingerprint expertise performance (measured by participants' performance on the xQ) over multiple training sessions, future research could extend the current experiment (e.g., by assessing novices over weeks or months of training, and incorporating multiple, varied training tasks) to further explore the benefit of interleaved practice. My thesis was one of the first studies to investigate the development of fingerprint expertise using interleaved practice and provides a platform for future research to build upon.

4.7 Conclusion

My thesis investigated the effect of a novel training paradigm on the fingerprint expertise performance of a small sample of novices. In incorporating the principle of

interleaved practice, my training tasks required participants to respond to either fingerprints from different fingers (mixing) or fingerprints from the same finger (massing). Despite the limited research combining interleaved practice and fingerprint examination, it was hypothesised that mixing fingerprint exemplars would lead to a significantly greater learning benefit on a test of fingerprint expertise performance (the xQ) than massing exemplars. However, the results suggested that there were no significant improvements in performance on the xQ over sessions, possibly due to an insufficient amount of training on a highly difficult and complex task. Massing exemplars did, however, lead to significant improvement on the training task itself, indicating that, while some learning did occur, it did not transfer to performance on the test. These results suggest that mixing and massing exemplars may have differential effects on learning in fingerprint examination, and that further studies should be conducted in the fascinating and rapidly growing field of perceptual expertise research.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*(1), 3-27. https://doi.org/10.1037/0033-2909.102.1.3

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387,* 401-406. https://doi.org/10.1038/387401a0

Bilalic, M., McLeod, P., & Gobet, F. (2009). Specialization effect and its influence on memory and problem solving in expert chess players. *Cognitive Science, 33,* 1117-1143. https://doi.org/10.1111/j.1551-6709.2009.01030.x

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances learning: The roles of discrimination and retrieval. *Memory & Cognition, 41,* 392-402. https://doi.org/10.3758/s13421-012-0272-7

Busey, T. A., & Parada, F. J. (2010). The nature of expertise in fingerprint examiners. *Psychonomic Bulletin & Review, 17*(2), 155-160. https://doi.org/10.3758/PBR.17.2.155

Busey, T., Swofford, H. J., Vanderkolk, J., & Emerick, B. (2015). The impact of fatigue on latent print examinations as revealed by behavioral and eye gaze testing. *Forensic Science International, 251,* 202-208. https://doi.org/10.1016/j.forsciint.2015..03.028

Busey. T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts, *Vision Research, 45,* 431-448. https://doi.org/10.1016/j.visres.2004.08.021

Campbell, A. (2011). *The fingerprint inquiry report.* Retrieved from http://www.thefingerprintinquiryscotland.org.uk/inquiry/3127-2.html

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition, 41,* 671-682. https://doi.org/10.3758/s13421-012-0291-4

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 & Cognition, 42,* 481-495. https://doi.org/10.3758/s1342-013-0371-0

Carvalho, P. F., & Goldstone, R. L. (2014b). Effects of interleaved and blocked study on delayed test of category learning generalisation. *Frontiers in Psychology, 5,* no. 936. https://doi.org/10.3389/fpsyg.2014.00936

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(11), 1699-1719. https://doi.org/10.1037/xlm0000406

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

Cole, S. A. (2005). More than zero: Accounting for error in latent fingerprint identification. *Journal of Criminal Law & Criminology, 95,* 985-1078.

de Croock, M. B. M., & van Merrienboer, J. J. G. (2007). Paradoxical effect of information presentation formats and contextual interference on transfer of a complex cognitive skill. *Computers in Human Behavior, 23,* 1740-1761.

Dosher, B. A., & Lu, Z-L. (2005). Perceptual learning in clear displays optimizes perceptual expertise: Learning the limiting process. *Proceedings of the National Academy of Sciences of the United States of America, 102*(14), 5286-5290. https://doi.org/10.1073/pnas.0500492102

Dror, I. E. (2016). A hierarchy of expert performance. *Journal of Applied Research in Memory and Cognition, 5,* 121-127.

Dror, I. E., & Charlton, D. (2006). Why experts make errors. *Journal of Forensic Identification, 56*(4), 600-616.

Dror, I. E., & Cole, S. A. (2010). The vision in "blind" justice: Expert perception, judgment, and visual cognition in forensic pattern recognition. *Psychonomic Bulletin & Review, 17*(2), 161-167. https://doi.org/10.3758/PBR.17.2.161

Dror, I. E., & Mnookin, J. L. (2010). The use of technology in human expert domains: Challenges and risks arising from the use of automated fingerprint identification systems in forensic science. *Law, Probability & Risk, 9,* 47-67.

Dunlosky, J., Rawson, K. A., Marsh. E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1). 4-58. https://doi.org/10.1177/1529100612453266

Dunphy, S. M., & Milbourne, C. C. (2009). Using word scrambles as an information systems creativity warm-up exercise. *Journal of Information Systems Education, 20*(1), 5-12.

Edmond, G., Tangen, J.M., Searston, R. A., & Dror, I. E. (2015). Contextual bias and cross-contamination in the forensic sciences: The corrosive implications for investigations, plea bargains, trials and appeals. *Law, Probability and Risk, 14,* 1-25. https://doi.org/10.1093/lpr/mgu018

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

Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004). An admissions OSCE: The multiple mini-interview. *Medical Education, 38,* 314-326. https://doi.org/10.1046/j.1365-2923.2004.01776.x

Faulds, H. (1880). On the skin-furrows of the hand. *Nature, 22,* 605.

Federal Bureau of Investigation. (2012). *30-year-old murder solved: Fingerprint technology played key role.* https://fbi.gov/news/stories/30-year-old-murder-solved

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature, 3*(2), 191-197.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble' experts: A framework for studying expert object recognition processes. *Vision Research, 38,* 2401-2428.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673-1682.

GitHub. (2015). *Common 7 letter words.* https://github.com/powerlanguage/word-lists/blob/master/common-7-letter-words.txt

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition, 24*(5), 608-628.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York, NY: Wiley.

Hall, K. G., Domingues, D. A., & Cavazos, R. (1994). Contextual interference effects with skilled baseball players. *Perceptual and Motor Skills, 78,* 835-841.

Hawkey, D. J. C., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature Neuroscience, 7*(1), 1055-1056.

Innocence Project. (2019). *Match in national fingerprint database establishes innocence of Baton Rouge man after 36 years in prison.* https://www.innocenceproject.org/fingerprint-database-match-establishes-archie-williams-innocence/

Jeter, P. E., Dosher, B. A., Lu, Z. L., & Petrov, A. P. (2009). Task precision at transfer determines specificity of perceptual learning. *Journal of Vision, 9*(3), 1-13.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26,* 97-103. https://doi.org/10.1002/acp.1801

Kellman, P. J., Mnookin, J. L., Erlikhman, G., Garrigan, P., Ghose, T., Mettler, E., Charlton, D., & Dror, I. E. (2014). Forensic comparison and matching of fingerprints: Using quantitative image measures for estimating error rates through understanding and predicting difficulty. *PLoS ONE, 9*(5), e94617. https://doi/org/10.1371/journal.pone.0094617

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

Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of Experimental Psychology, 51,* 239-243.

Lakens, D., Scheel., A. M., & Isager, P. M (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259-269. https://doi.org/10.1177/2515245918770963

Lin, C-H., Knowlton, B. J., Chiang, M-C., Iacoboni, M., Udompholkul, P., & Wu, A. D. (2011). Brain-behaviour correlates of optimising through interleaved practice. *NeuroImage, 56,* 1758-1772.

Liu, Z. L., & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision Research, 40*(1), 97-109.

Lu, Z. L., & Dosher, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America A, 16*(3). 764-778.

Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behaviour Analysis, 35,* 105-123.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 596-606.

Mielicki, M. (2019). *Exploring the necessary conditions for observing interleaved practice benefits in problem solving* [Doctoral dissertation, University of Illinois at Chicago]. ProQuest. https://search.proquest.com/openview/7919a0de4786bee39949cb579ca10357/1?pq-origsite=gscholar&cbl=18750&diss=y

Murman, D. L. (2015). The impact of age on cognition. *Seminars in Hearing, 36*(3), 111-121. https://doi.org/10.1055/s-0035-1555115

Mustonen, V., & Himberg, K. (2011). A novel approach to the education of fingerprint experts. *Forensic Science Policy and Management, 2,* 28-35. https://doi.org/10.1080/19409044.2011.564270

National Institute of Standards and Technology. (2012). *Latent print examination and human factors: Improving the practice through a systems approach: The report of the expert Working Group on human factors in latent print analysis.* Gaithersburg, MD: U.S. Department of Commerce, National Institute of Standards and Technology.

National Research Council, Committee on Identifying the Needs of the Forensic Science Community (NAS). (2009). *Strengthening forensic science in the United States: A path forward.* The National Academies Press.

Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual-learning in visual hyperacuity. *Science, 256*(5059), 1018-1021.

President's Council of Advisors on Science and Technology (PCAST). (2016). *Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods.* Executive Office of the President of the United States.

Rau, M. A., Aleven, V., & Rummel, N. (2013). Interleaved practice in multi-dimensional learning tasks: Which dimension should we interleave? *Learning and Instruction, 23,* 98-114. https://doi.org/10.1016/j.learninstruc.2012.07.003

Rinck, M., & Becker, E. S. (2005). A comparison of attentional biases and memory biases in women with social phobia and major depression. *Journal of Abnormal Psychology, 114*(1), 62-74.

Roads, B., Mozer, M. C., & Busey, T. A. (2016). Using highlighting to train attentional expertise. *PLoS ONE, 11*(1), e0146266. https://doi.org/10.1371/journal.pone.0146266

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review, 24,* 355-367. https://doi.org/10.1007/s10648-012-9201-3

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science, 35,* 481-498.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382-439.

Rosenthal, R. (1966). *Experimenter Effects in Behavioral Research.* East Norwalk, CT: Appleton-Century-Crofts.

Sagi, D. (2011). Perceptual learning in Vision Research. *Vision Research, 51*(13), 1552-1566.

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). The style of a stranger: Identification expertise generalizes to coarser level categories. *Psychonomic Bulletin & Review, 24*(4), 1324-1329. https://doi.org/10.3758/s13423-016-1211-6

Searston, R. A., & Tangen, J. M. (2017b). Expertise with unfamiliar objects is flexible to changes in task but not changes in class. *PLoS ONE, 12*(6), e0178403. https://doi.org/10.1371/journal.pone.0178403

Searston, R. A., & Tangen, J. M. (2017c). Training perceptual experts: Feedback, labels, and contrasts. *Canadian Journal of Experimental Psychology, 71*(1), 32-39. https://doi.org/10.1037/cep0000124

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. https://doi/org/10.1016/j.jarmac.2017.08.006

Searston, R. A., Tangen, J. M., & Eva, K. W. (2016). Putting bias into context: The role of familiarity in identification. *Law and Human Behavior, 40*(1), 50-64. https://doi.org/10.1037/lhb0000154

Searston, R. A., Thompson, M. B., Robson, S. G., Corbett, B. J., Ribeiro, G., Edmond, G., & Tangen, J. M. (2019). Truth and transparency in expertise research. *Journal of Expertise, 2*(4), 199-209.

Shen, J., Mack, M. L., & Palmeri, T. J. (2014). Studying real-world perceptual expertise. *Frontiers in Psychology, 5*, article no. 857.

Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics, 52*(5), 582-588.

Sibley, D. A. (2002). *Sibley's birding basics.* Toledo, Spain: Knopf.

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10*(2), 176-199. https://doi.org/10.1177/1745691615569000

Sorensen, L. J., & Woltz, D. J. (2016). Blocking as a friend of induction in verbal category learning. *Memory & Cognition, 44,* 1000-1013. https://doi.org/10.3758/s13421-016-0615-x

Stevenage, S. V., & Pitfield, C. (2016). Fact or friction: Examination of the transparency, reliability and sufficiency of the ACE-V method of fingerprint analysis. *Forensic Science International, 267,* 145-156.

Szpiro, S. F. A., Wright, B. A., & Carrasco, M. (2014). Learning one task by interleaving practice with another task. *Vision Research, 101,* 118-124. https://doi.org/10.1016/j.visres.2014.06.004

Tanaka, J. W. (2001). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General, 130,* 534-543.

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., Kent, K. M., & Searston, R. A. (2020). Collective intelligence in fingerprint analysis. *Cognitive Research: Principles and Implications, 5,* 23. https://doi.org/10.1186/s41235-020-00223-8

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457-482.

Tangen, J. M., Thompson, M. B., & McCarthy, D. J. (2011). Identifying fingerprint expertise. *Psychological Science, 22*(8), 995-997. https://doi.org/10.1177/0956797611414729

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24,* 837-484. https://doi.org/10.1002/acp.1598

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), e114759. https://doi.org/10.1371/journal.pone.0114759

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2013). Expertise in fingerprint identification. *Journal of Forensic Science, 58*(6), https://doi.org/10.1111/1556-4029.12203

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2014). Human matching performance of genuine crime scene latent fingerprints. *Law and Human Behavior, 38*(1), 84-93. https://doi.org/10.1037/lhb0000051

Tigner, R. B. (1999). Putting memory research to good use. *College Teaching, 47,* 149-152.

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,* 199-208.

Tran, T. B., Siemer, M., & Joormann, J. (2011). Implicit interpretation biases affect emotional vulnerability: A training study. *Cognition and Emotion, 25*(3), 546-558.

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 of the United States of America, 108*(19), 7733-7738. https://doi.org/10.1073/pnas.1018707108

Vokey, J. R., Tangen, J. M., & Cole, S. A. (2009). On the preliminary psychophysics of fingerprint identification. *Quarterly Journal of Experimental Psychology, 62,* 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 & Cognition, 39,* 750-763. https://doi.org/10.3758/s13421-010-0063-y

Westfall, J. (2020). *PANGEA: Power ANalysis for GEneral Anova designs*. PANGEA. https://jakewestfall.shinyapps.io/pangea/

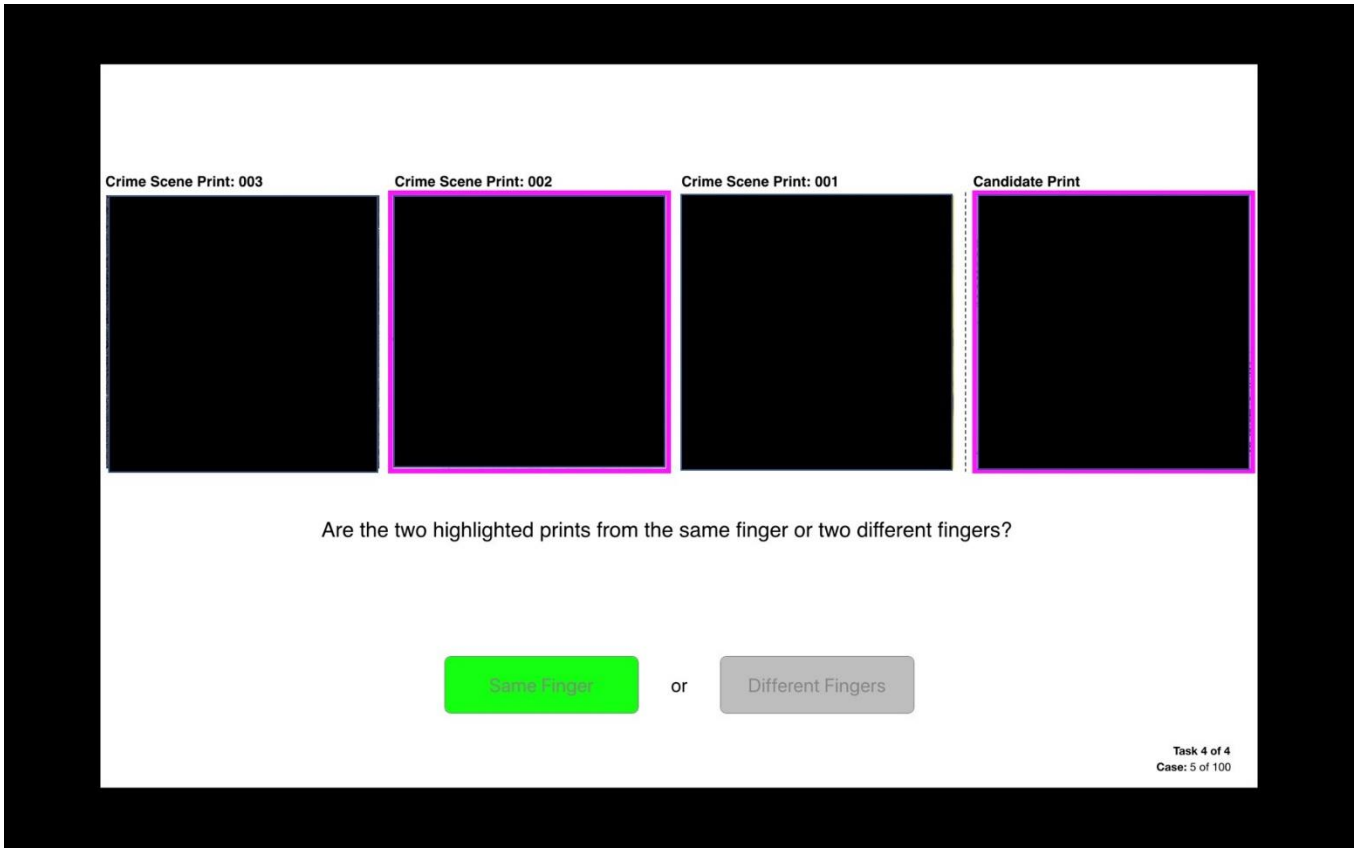
White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015). Error rates in users of automatic face recognition software. *PLoS ONE, 10*(10), 1-14. https://doi.org/10.1371/journal.pone.0139827

Wright, B. A., & Zhang, Y. (2009). A review of the generalization of auditory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1515). 301-311.

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition, 41,* 16-27. https://doi.org/10.3758/s13421-012-0238-9

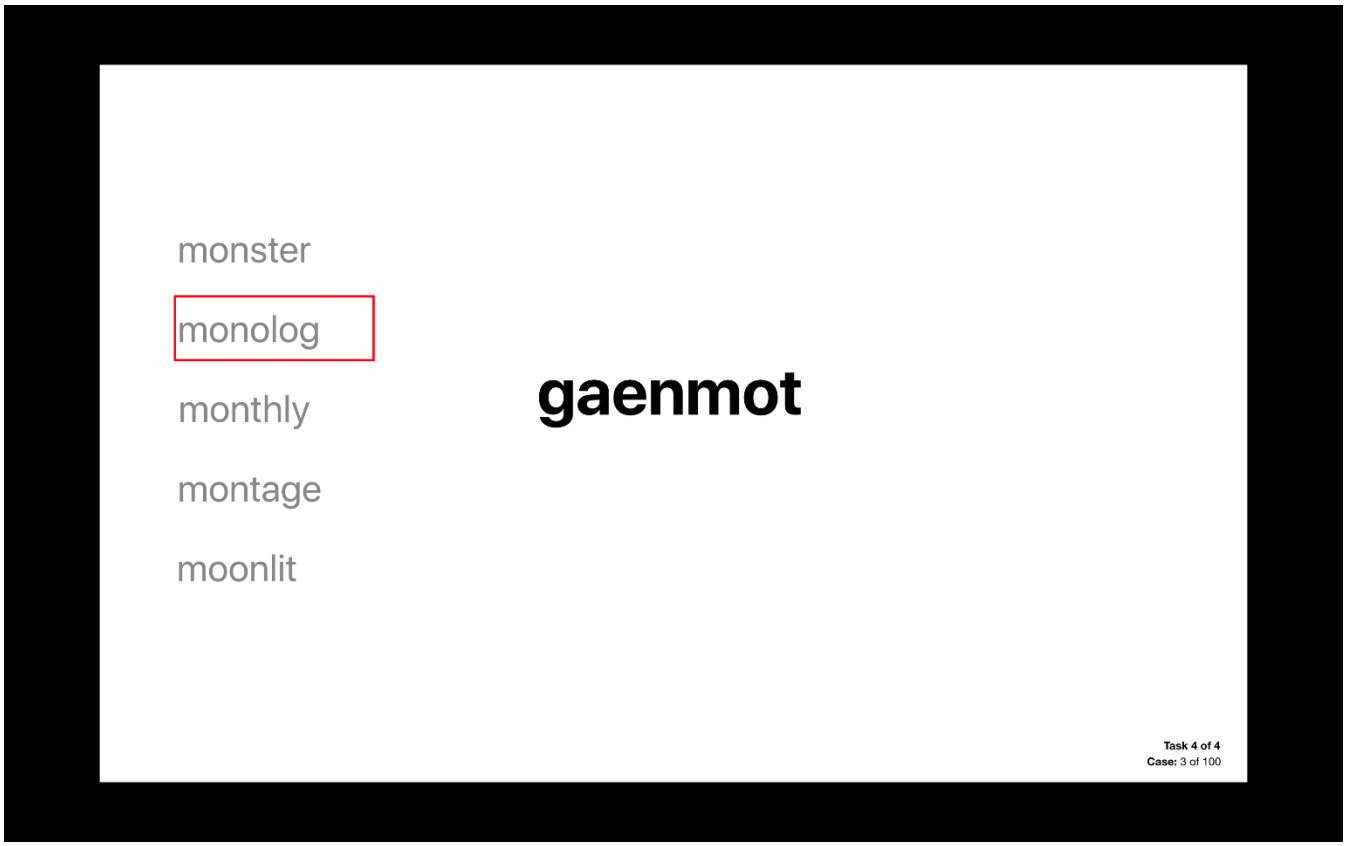
Appendix A

Example of a trial on the fingerprint training tasks (fingerprints blocked out due to legal and ethical constraints)



Appendix B

Example of a trial on the Control training task



Appendix C

ANOVA output excluding my own data, produced in R Studio. Pseudonym blocked out for confidentiality reasons.

```r
```{r, messages = FALSE}
mydata <- read.csv("finaldata.csv", header = TRUE)
mydata <- mydata %>% filter(!Participant == "[redacted]")

## Specify Variables
mydata$Participant <- factor(mydata$Participant)
mydata$Condition <- factor(mydata$Condition,levels = c("Control","Mixed","Massed"))
mydata$Session = factor(mydata$Session, levels = c("1","2","3","4","5","6","7","8","9","10","11"),
ordered = TRUE)
```
```

ANOVA

```r
```{r}
ezANOVA(data=mydata, dv=xQ, wid= Participant, between = Condition, within = Session)
```
```

R Console | data.frame 3 x 7 | data.frame 2 x 4 | data.frame 2 x 7

	Effect <chr>	DFn <dbl>	DFd <dbl>	F <dbl>	p <dbl>	p<.05 <chr>	ges <dbl>
2	Condition	2	15	2.1340772	0.1528916		0.07032462
3	Session	10	150	1.3818736	0.1937030		0.06334946
4	Condition:Session	20	150	0.5168672	0.9562549		0.04815825

R Console | data.frame 3 x 7 | data.frame 2 x 4 | data.frame 2 x 7

	Effect <chr>	W <dbl>	p <dbl>	p<.05 <chr>
3	Session	0.01256832	0.7036058	
4	Condition:Session	0.01256832	0.7036058	

Treatment-control contrasts

```r
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
-0.188095 -0.052315  0.003889  0.050463  0.188889

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                0.4163780  0.0096775  43.025  < 2e-16 ***
ConditionMixed             0.0093795  0.0142449   0.658 0.511168
ConditionMassed            0.0510966  0.0149923   3.408 0.000822 ***
Session.L                  0.0587213  0.0320967   1.830 0.069128 .
Session.Q                 -0.0291541  0.0320967  -0.908 0.365034
Session.C                 -0.0407618  0.0320967  -1.270 0.205883
Session^4                 -0.0056316  0.0320967  -0.175 0.860935
Session^5                 -0.0012074  0.0320967  -0.038 0.970037
Session^6                  0.0016259  0.0320967   0.051 0.959659
Session^7                 -0.0088442  0.0320967  -0.276 0.783240
Session^8                  0.0366364  0.0320967   1.141 0.255341
Session^9                  0.0325718  0.0320967   1.015 0.311685
Session^10                 0.0006629  0.0320967   0.021 0.983548
ConditionMixed:Session.L   0.0087276  0.0472450   0.185 0.853668
ConditionMassed:Session.L -0.0380630  0.0497239  -0.765 0.445075
ConditionMixed:Session.Q   0.0691099  0.0472450   1.463 0.145425
ConditionMassed:Session.Q  0.0157638  0.0497239   0.317 0.751624
ConditionMixed:Session.C   0.0184824  0.0472450   0.391 0.696152
ConditionMassed:Session.C  0.0128561  0.0497239   0.259 0.796305
ConditionMixed:Session^4   0.0178959  0.0472450   0.379 0.705332
ConditionMassed:Session^4  0.0134502  0.0497239   0.270 0.787116
ConditionMixed:Session^5  -0.0155466  0.0472450  -0.329 0.742524
ConditionMassed:Session^5  0.0246039  0.0497239   0.495 0.621392
ConditionMixed:Session^6  -0.0026923  0.0472450  -0.057 0.954626
ConditionMassed:Session^6  0.0564759  0.0497239   1.136 0.257692
ConditionMixed:Session^7  -0.0217640  0.0472450  -0.461 0.645647
ConditionMassed:Session^7  0.0014662  0.0497239   0.029 0.976512
ConditionMixed:Session^8  -0.0487082  0.0472450  -1.031 0.304063
ConditionMassed:Session^8 -0.0524570  0.0497239  -1.055 0.292983
ConditionMixed:Session^9  -0.0316329  0.0472450  -0.670 0.504081
ConditionMassed:Session^9 -0.0261828  0.0497239  -0.527 0.599203
ConditionMixed:Session^10 -0.0232728  0.0472450  -0.493 0.622951
ConditionMassed:Session^10 -0.0128976  0.0497239  -0.259 0.795663
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Helmert contrasts

```r
options(contrasts=c("contr.helmert","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
-0.188095 -0.052315  0.003889  0.050463  0.188889

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                0.4365368  0.0060922  71.655  < 2e-16 ***
Condition1                 0.0046898  0.0071225   0.658 0.511168
Condition2                 0.0154690  0.0044950   3.441 0.000733 ***
Session.L                  0.0489428  0.0202055   2.422 0.016508 *
Session.Q                 -0.0008628  0.0202055  -0.043 0.965990
Session.C                 -0.0303157  0.0202055  -1.500 0.135431
Session^4                  0.0048171  0.0202055   0.238 0.811864
Session^5                  0.0018117  0.0202055   0.090 0.928665
Session^6                  0.0195538  0.0202055   0.968 0.334587
Session^7                 -0.0156101  0.0202055  -0.773 0.440883
Session^8                  0.0029146  0.0202055   0.144 0.885479
Session^9                  0.0132999  0.0202055   0.658 0.511307
Session^10                -0.0113939  0.0202055  -0.564 0.573587
Condition1:Session.L       0.0043638  0.0236225   0.185 0.853668
Condition2:Session.L      -0.0141423  0.0149082  -0.949 0.344200
Condition1:Session.Q       0.0345549  0.0236225   1.463 0.145425
Condition2:Session.Q      -0.0062637  0.0149082  -0.420 0.674922
Condition1:Session.C       0.0092412  0.0236225   0.391 0.696152
Condition2:Session.C       0.0012050  0.0149082   0.081 0.935678
Condition1:Session^4       0.0089479  0.0236225   0.379 0.705332
Condition2:Session^4       0.0015007  0.0149082   0.101 0.919938
Condition1:Session^5      -0.0077733  0.0236225  -0.329 0.742524
Condition2:Session^5       0.0107924  0.0149082   0.724 0.470138
Condition1:Session^6      -0.0013461  0.0236225  -0.057 0.954626
Condition2:Session^6       0.0192740  0.0149082   1.293 0.197872
Condition1:Session^7      -0.0108820  0.0236225  -0.461 0.645647
Condition2:Session^7       0.0041161  0.0149082   0.276 0.782822
Condition1:Session^8      -0.0243541  0.0236225  -1.031 0.304063
Condition2:Session^8      -0.0093676  0.0149082  -0.628 0.530641
Condition1:Session^9      -0.0158164  0.0236225  -0.670 0.504081
Condition2:Session^9      -0.0034555  0.0149082  -0.232 0.816994
Condition1:Session^10     -0.0116364  0.0236225  -0.493 0.622951
Condition2:Session^10     -0.0004204  0.0149082  -0.028 0.977538
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

ANOVA output excluding the data of the extra participant, produced in R Studio. Pseudonym blocked out for confidentiality reasons.

```r
mydata <- read.csv("finaldata.csv", header = TRUE)
mydata <- mydata %>% filter(!Participant == "[redacted]")

## Specify Variables
mydata$Participant <- factor(mydata$Participant)
mydata$Condition <- factor(mydata$Condition,levels = c("Control","Mixed","Massed"))
mydata$Session = factor(mydata$Session, levels = c("1","2","3","4","5","6","7","8","9","10","11"),
ordered = TRUE)
```

ANOVA

```r
ezANOVA(data=mydata, dv=xQ, wid= Participant, between = Condition, within = Session)
```

R Console | data.frame 3 x 7 | data.frame 2 x 4 | data.frame 2 x 7

	Effect <chr>	DFn <dbl>	DFd <dbl>	F <dbl>	p <dbl>	p<.05 <chr>	ges <dbl>
2	Condition	2	15	0.6947564	0.5145637		0.02802808
3	Session	10	150	1.4240732	0.1747143		0.06137190
4	Condition:Session	20	150	0.5311019	0.9495751		0.04650197

R Console | data.frame 3 x 7 | data.frame 2 x 4 | data.frame 2 x 7

	Effect <chr>	W <dbl>	p <dbl>	p<.05 <chr>
3	Session	0.005989058	0.3934066	
4	Condition:Session	0.005989058	0.3934066	

Treatment-control contrasts

```r
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
-0.188889 -0.053472  0.007407  0.050695  0.186111

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                 0.419192   0.010891  38.489   <2e-16 ***
ConditionMixed              0.006566   0.015403   0.426   0.6705
ConditionMassed             0.031818   0.015403   2.066   0.0404 *
Session.L                   0.053853   0.036122   1.491   0.1379
Session.Q                  -0.013087   0.036122  -0.362   0.7176
Session.C                  -0.046509   0.036122  -1.288   0.1997
Session^4                  -0.000657   0.036122  -0.018   0.9855
Session^5                  -0.021721   0.036122  -0.601   0.5485
Session^6                  -0.004012   0.036122  -0.111   0.9117
Session^7                  -0.019893   0.036122  -0.551   0.5826
Session^8                   0.041861   0.036122   1.159   0.2482
Session^9                   0.041531   0.036122   1.150   0.2519
Session^10                  0.001794   0.036122   0.050   0.9604
ConditionMixed:Session.L    0.013596   0.051085   0.266   0.7905
ConditionMassed:Session.L  -0.025161   0.051085  -0.493   0.6230
ConditionMixed:Session.Q    0.053043   0.051085   1.038   0.3006
ConditionMassed:Session.Q  -0.013435   0.051085  -0.263   0.7929
ConditionMixed:Session.C    0.024230   0.051085   0.474   0.6359
ConditionMassed:Session.C   0.036402   0.051085   0.713   0.4771
ConditionMixed:Session^4    0.012921   0.051085   0.253   0.8006
ConditionMassed:Session^4   0.005640   0.051085   0.110   0.9122
ConditionMixed:Session^5    0.004967   0.051085   0.097   0.9227
ConditionMassed:Session^5   0.020535   0.051085   0.402   0.6882
ConditionMixed:Session^6    0.002946   0.051085   0.058   0.9541
ConditionMassed:Session^6   0.055184   0.051085   1.080   0.2816
ConditionMixed:Session^7   -0.010716   0.051085  -0.210   0.8341
ConditionMassed:Session^7  -0.001580   0.051085  -0.031   0.9754
ConditionMixed:Session^8   -0.053932   0.051085  -1.056   0.2926
ConditionMassed:Session^8  -0.076840   0.051085  -1.504   0.1345
ConditionMixed:Session^9   -0.040592   0.051085  -0.795   0.4280
ConditionMassed:Session^9  -0.023465   0.051085  -0.459   0.6466
ConditionMixed:Session^10  -0.024404   0.051085  -0.478   0.6335
ConditionMassed:Session^10 -0.006745   0.051085  -0.132   0.8951
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Helmert contrasts

```r
options(contrasts=c("contr.helmert","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
-0.188889 -0.053472  0.007407  0.050695  0.186111

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                 0.4319865  0.0062881  68.699   <2e-16 ***
Condition1                  0.0032828  0.0077013   0.426   0.6705
Condition2                  0.0095118  0.0044464   2.139   0.0339 *
Session.L                   0.0499980  0.0208553   2.397   0.0176 *
Session.Q                   0.0001159  0.0208553   0.006   0.9956
Session.C                  -0.0262988  0.0208553  -1.261   0.2091
Session^4                   0.0055299  0.0208553   0.265   0.7912
Session^5                  -0.0132205  0.0208553  -0.634   0.5270
Session^6                   0.0153644  0.0208553   0.737   0.4623
Session^7                  -0.0239910  0.0208553  -1.150   0.2517
Session^8                  -0.0017302  0.0208553  -0.083   0.9340
Session^9                   0.0201787  0.0208553   0.968   0.3347
Session^10                 -0.0085886  0.0208553  -0.412   0.6810
Condition1:Session.L        0.0067979  0.0255424   0.266   0.7905
Condition2:Session.L       -0.0106529  0.0147469  -0.722   0.4711
Condition1:Session.Q        0.0265213  0.0255424   1.038   0.3006
Condition2:Session.Q       -0.0133186  0.0147469  -0.903   0.3678
Condition1:Session.C        0.0121149  0.0255424   0.474   0.6359
Condition2:Session.C        0.0080956  0.0147469   0.549   0.5838
Condition1:Session^4        0.0064606  0.0255424   0.253   0.8006
Condition2:Session^4       -0.0002737  0.0147469  -0.019   0.9852
Condition1:Session^5        0.0024836  0.0255424   0.097   0.9227
Condition2:Session^5        0.0060172  0.0147469   0.408   0.6838
Condition1:Session^6        0.0014730  0.0255424   0.058   0.9541
Condition2:Session^6        0.0179037  0.0147469   1.214   0.2265
Condition1:Session^7       -0.0053579  0.0255424  -0.210   0.8341
Condition2:Session^7        0.0012593  0.0147469   0.085   0.9320
Condition1:Session^8       -0.0269662  0.0255424  -1.056   0.2926
Condition2:Session^8       -0.0166245  0.0147469  -1.127   0.2612
Condition1:Session^9       -0.0202961  0.0255424  -0.795   0.4280
Condition2:Session^9       -0.0010563  0.0147469  -0.072   0.9430
Condition1:Session^10      -0.0122021  0.0255424  -0.478   0.6335
Condition2:Session^10       0.0018192  0.0147469   0.123   0.9020
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix D

Participant information sheet

Participant Information

THE UNIVERSITY of ADELAIDE

PROJECT: Improving learning in fingerprint examination

RESEARCHER: Claire Walker, School of Psychology, University of Adelaide

SUPERVISOR: Dr Rachel Searston, School of Psychology, University of Adelaide

PARTICIPATION: How do novices in fingerprint matching become experts? What is the most effective way to train these experts? You are invited to participate in a study seeking to answer these questions. This study will involve 11 sessions of approximately 20 to 30 minutes each on 11 consecutive days (with the exception of day 11, which will run for 12 minutes). This research could potentially be useful for forensic experts seeking to enhance the quality of fingerprint training, and, more broadly, in the field of perceptual expertise itself. There are no foreseeable risks outside those that exist in everyday life. At the conclusion of your participation, the researcher will disclose the study's aims, methodology and variables of interest, and answer any questions you may have. The researcher will also be available for assistance during your participation.

CONSENT: Your participation is completely voluntary, and you are free to withdraw from the study at any time without consequence. You will receive a $20 Coles/Myer gift card for every hour of participation 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 for the hours of participation you have completed.

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 by a unique code upon completion. You will not be identifiable by this code, but your performance in this experiment will be recorded. 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 20/31). 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: 20/31

Appendix E

Participant consent form

Consent Form

THE UNIVERSITY of ADELAIDE

PROJECT: Improving learning in fingerprint examination

RESEARCHER: Claire Walker, School of Psychology, University of Adelaide

SUPERVISOR: Dr Rachel Searston, School of Psychology, University of Adelaide

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 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
   (b) my participation in the research will have no effect on my academic grades, enrolment or future employment.

Name of participant: _____

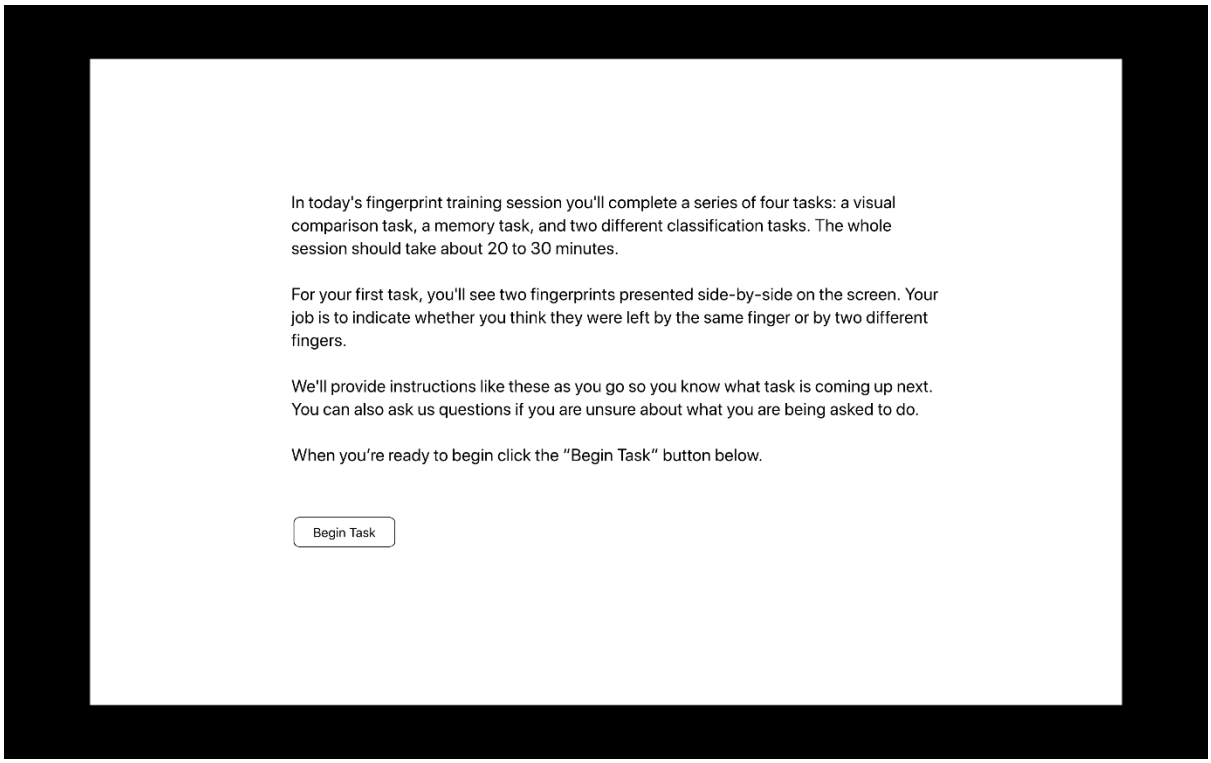
Signature: _____ Date: _____

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

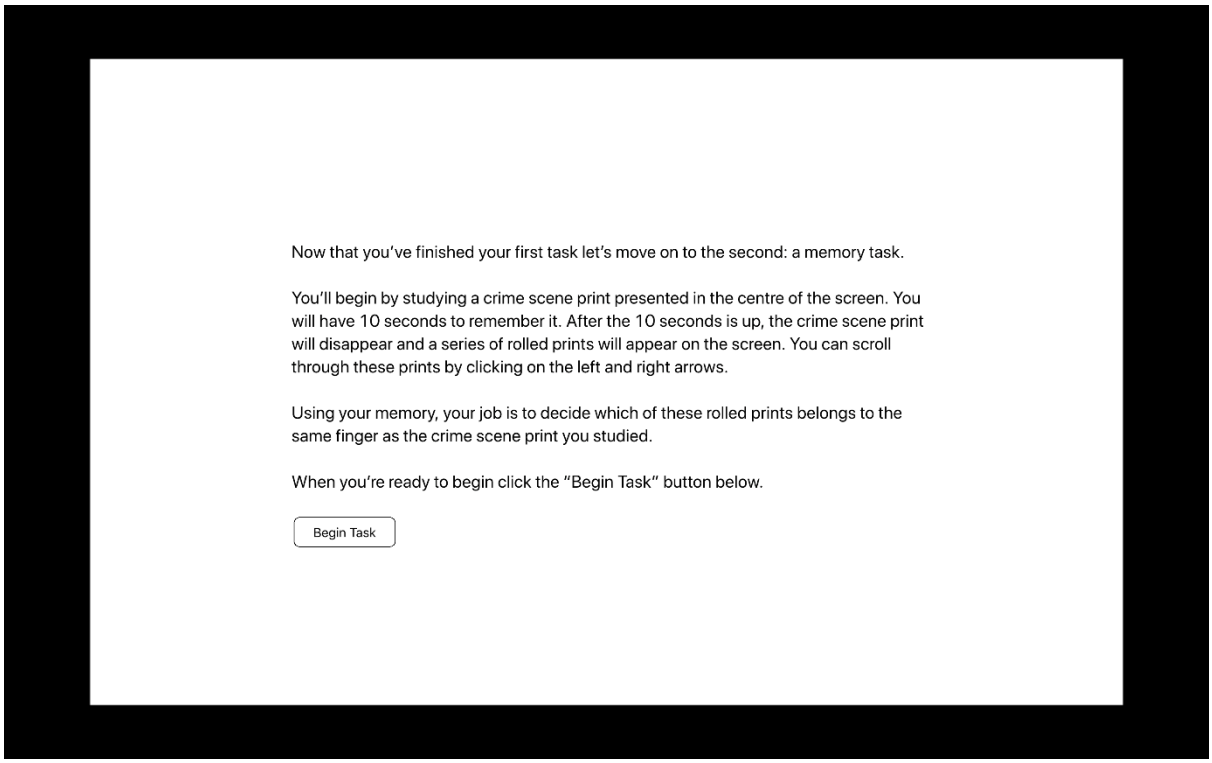
Project Number: 20/31

Appendix F

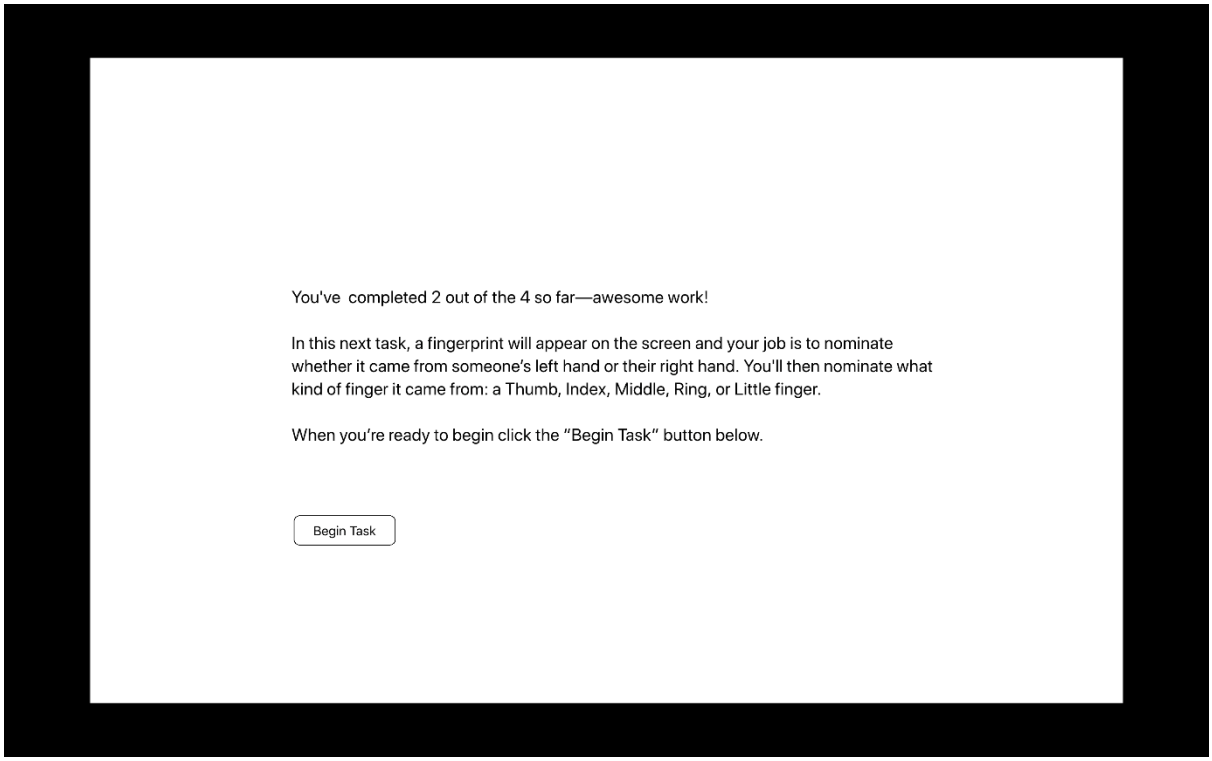
Initial instructions, prior to task 1



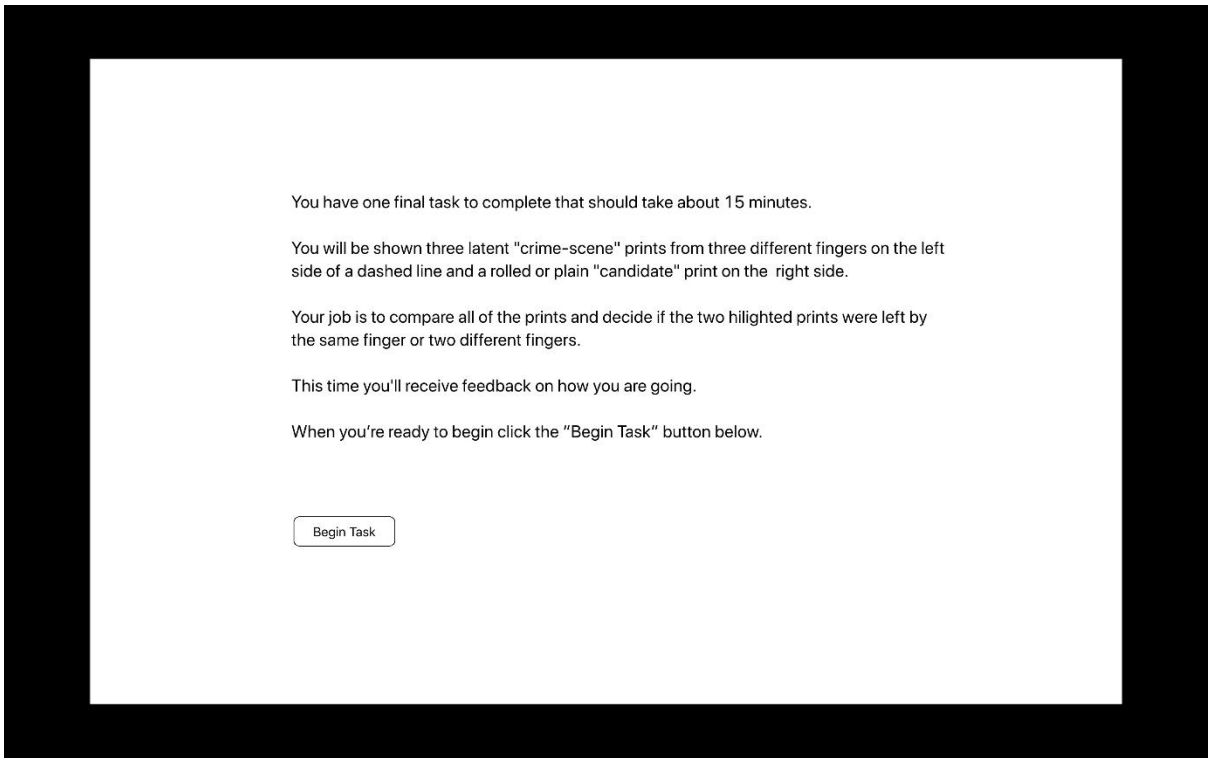
Task 2 instructions



Task 3 instructions

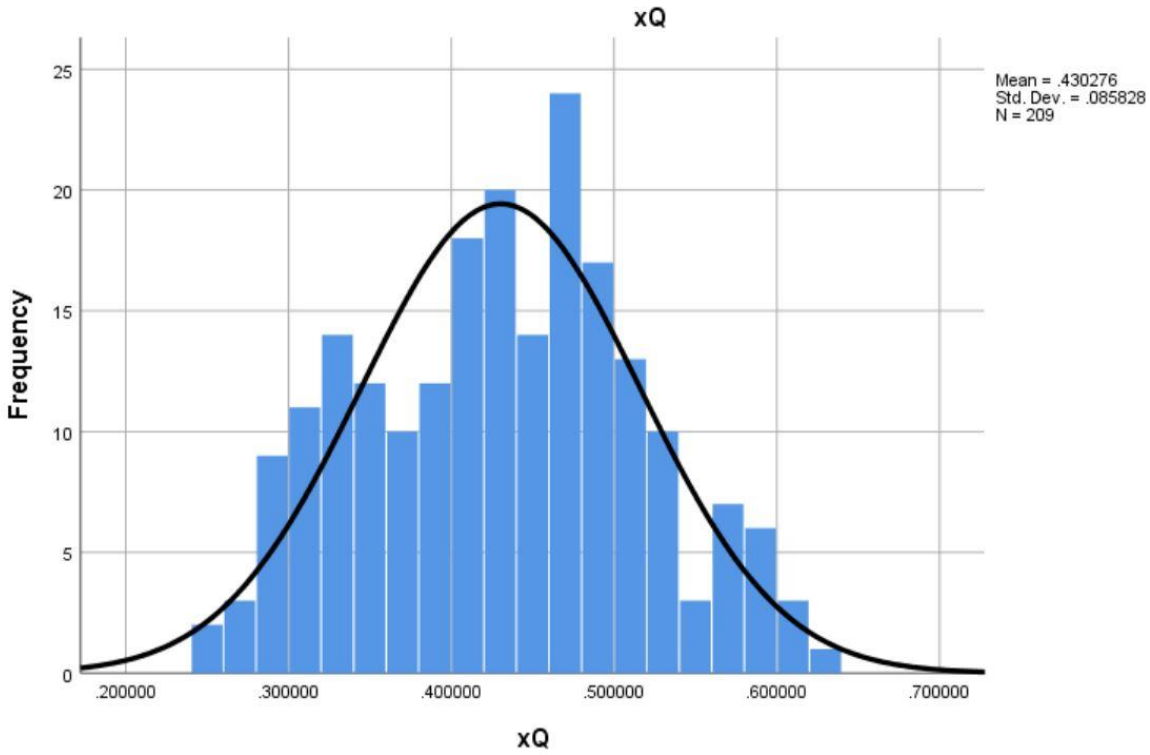


Task 4 instructions



Appendix G

Histogram of the xQ data, produced in SPSS



Appendix H

Full ANOVA output for performance on the xQ, produced in R Studio

```r
ezANOVA(data=mydata, dv=xQ, wid= Participant, between = Condition, within = Session)
```

R Console | data.frame 3 x 7 | data.frame 2 x 4 | data.frame 2 x 7

	Effect	DFn	DFd	F	p	p<.05	ges
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>
2	Condition	2	16	0.8791711	0.4342507		0.03224406
3	Session	10	160	1.4626098	0.1580007		0.05988397
4	Condition:Session	20	160	0.5486014	0.9408883		0.04560532

	Effect	W	p	p<.05
	<chr>	<dbl>	<dbl>	<chr>
3	Session	0.01370961	0.5704893	
4	Condition:Session	0.01370961	0.5704893	

Treatment-control contrasts

```r
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
-0.188095 -0.053704  0.001852  0.049074  0.188889

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                4.164e-01  9.938e-03  41.898   <2e-16 ***
ConditionMixed             9.380e-03  1.463e-02   0.641   0.5222
ConditionMassed            3.463e-02  1.463e-02   2.367   0.0190 *
Session.L                  5.872e-02  3.296e-02   1.782   0.0765 .
Session.Q                 -2.915e-02  3.296e-02  -0.885   0.3776
Session.C                 -4.076e-02  3.296e-02  -1.237   0.2178
Session^4                 -5.632e-03  3.296e-02  -0.171   0.8645
Session^5                 -1.207e-03  3.296e-02  -0.037   0.9708
Session^6                  1.626e-03  3.296e-02   0.049   0.9607
Session^7                 -8.844e-03  3.296e-02  -0.268   0.7888
Session^8                  3.664e-02  3.296e-02   1.112   0.2679
Session^9                  3.257e-02  3.296e-02   0.988   0.3244
Session^10                 6.629e-04  3.296e-02   0.020   0.9840
ConditionMixed:Session.L   8.728e-03  4.852e-02   0.180   0.8574
ConditionMassed:Session.L -3.003e-02  4.852e-02  -0.619   0.5367
ConditionMixed:Session.Q   6.911e-02  4.852e-02   1.424   0.1561
ConditionMassed:Session.Q  2.633e-03  4.852e-02   0.054   0.9568
ConditionMixed:Session.C   1.848e-02  4.852e-02   0.381   0.7037
ConditionMassed:Session.C  3.065e-02  4.852e-02   0.632   0.5283
ConditionMixed:Session^4   1.790e-02  4.852e-02   0.369   0.7127
ConditionMassed:Session^4  1.061e-02  4.852e-02   0.219   0.8271
ConditionMixed:Session^5  -1.555e-02  4.852e-02  -0.320   0.7490
ConditionMassed:Session^5  2.133e-05  4.852e-02   0.000   0.9996
ConditionMixed:Session^6  -2.692e-03  4.852e-02  -0.055   0.9558
ConditionMassed:Session^6  4.955e-02  4.852e-02   1.021   0.3085
ConditionMixed:Session^7  -2.176e-02  4.852e-02  -0.449   0.6543
ConditionMassed:Session^7 -1.263e-02  4.852e-02  -0.260   0.7949
ConditionMixed:Session^8  -4.871e-02  4.852e-02  -1.004   0.3168
ConditionMassed:Session^8 -7.162e-02  4.852e-02  -1.476   0.1417
ConditionMixed:Session^9  -3.163e-02  4.852e-02  -0.652   0.5152
ConditionMassed:Session^9 -1.451e-02  4.852e-02  -0.299   0.7653
ConditionMixed:Session^10 -2.327e-02  4.852e-02  -0.480   0.6320
ConditionMassed:Session^10 -5.613e-03  4.852e-02  -0.116   0.9080
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Helmert contrasts

```r
options(contrasts=c("contr.helmert","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
-0.188095 -0.053704  0.001852  0.049074  0.188889

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                0.4310485  0.0060480  71.271   <2e-16 ***
Condition1                 0.0046898  0.0073141   0.641   0.5222
Condition2                 0.0099807  0.0043297   2.305   0.0223 *
Session.L                  0.0516207  0.0200589   2.573   0.0109 *
Session.Q                 -0.0052399  0.0200589  -0.261   0.7942
Session.C                 -0.0243830  0.0200589  -1.216   0.2258
Session^4                  0.0038717  0.0200589   0.193   0.8472
Session^5                 -0.0063825  0.0200589  -0.318   0.7507
Session^6                  0.0172438  0.0200589   0.860   0.3911
Session^7                 -0.0203082  0.0200589  -1.012   0.3127
Session^8                 -0.0034715  0.0200589  -0.173   0.8628
Session^9                  0.0171923  0.0200589   0.857   0.3926
Session^10                -0.0089658  0.0200589  -0.447   0.6554
Condition1:Session.L       0.0043638  0.0242580   0.180   0.8574
Condition2:Session.L      -0.0114643  0.0143600  -0.798   0.4257
Condition1:Session.Q       0.0345549  0.0242580   1.424   0.1561
Condition2:Session.Q      -0.0106407  0.0143600  -0.741   0.4597
Condition1:Session.C       0.0092412  0.0242580   0.381   0.7037
Condition2:Session.C       0.0071377  0.0143600   0.497   0.6198
Condition1:Session^4       0.0089479  0.0242580   0.369   0.7127
Condition2:Session^4       0.0005554  0.0143600   0.039   0.9692
Condition1:Session^5      -0.0077733  0.0242580  -0.320   0.7490
Condition2:Session^5       0.0025982  0.0143600   0.181   0.8566
Condition1:Session^6      -0.0013461  0.0242580  -0.055   0.9558
Condition2:Session^6       0.0169640  0.0143600   1.181   0.2391
Condition1:Session^7      -0.0108820  0.0242580  -0.449   0.6543
Condition2:Session^7      -0.0005820  0.0143600  -0.041   0.9677
Condition1:Session^8      -0.0243541  0.0242580  -1.004   0.3168
Condition2:Session^8      -0.0157538  0.0143600  -1.097   0.2741
Condition1:Session^9      -0.0158164  0.0242580  -0.652   0.5152
Condition2:Session^9       0.0004369  0.0143600   0.030   0.9758
Condition1:Session^10     -0.0116364  0.0242580  -0.480   0.6320
Condition2:Session^10      0.0020078  0.0143600   0.140   0.8890
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix I

Full ANOVA output for performance on the fingerprint training tasks, produced in R Studio

```r
ezANOVA(data=mydata_Training, dv=Training, wid= Participant, between = Condition, within = Session)
```

R Console | data.frame 3 x 7 | data.frame 2 x 4 | data.frame 2 x 7

	Effect	DFn	DFd	F	p	p<.05	ges
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>
2	Condition	1	10	6.664019	0.027346962	*	0.28643816
3	Session	9	90	1.584638	0.131802971		0.05927512
4	Condition:Session	9	90	2.919181	0.004429503	*	0.10400329

```r
ezANOVA(data=mydata, dv=xQ, wid= Participant, between = Condition, within = Session)
```

R Console | data.frame 3 x 7 | data.frame 2 x 4 | data.frame 2 x 7

	Effect	W	p	p<.05
	<chr>	<dbl>	<dbl>	<chr>
3	Session	0.01370961	0.5704893	
4	Condition:Session	0.01370961	0.5704893	

Treatment-control contrasts

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

```
Call:
lm(formula = Training ~ Condition * Session, data = mydata_Training)

Residuals:
      Min        1Q    Median        3Q       Max
-0.198333 -0.043333 -0.001667  0.045417  0.166667

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               0.581500   0.009654  60.235  < 2e-16 ***
ConditionMassed           0.086500   0.013653   6.336 6.82e-09 ***
Session.L                -0.013670   0.030528  -0.448  0.65527
Session.Q                -0.013346   0.030528  -0.437  0.66293
Session.C                -0.015150   0.030528  -0.496  0.62079
Session^4                 0.006295   0.030528   0.206  0.83704
Session^5                -0.009071   0.030528  -0.297  0.76699
Session^6                 0.013364   0.030528   0.438  0.66250
Session^7                 0.010461   0.030528   0.343  0.73258
Session^8                 0.006607   0.030528   0.216  0.82910
Session^9                 0.008753   0.030528   0.287  0.77492
ConditionMassed:Session.L  0.117344   0.043173   2.718  0.00774 **
ConditionMassed:Session.Q -0.017118   0.043173  -0.396  0.69259
ConditionMassed:Session.C  0.035662   0.043173   0.826  0.41076
ConditionMassed:Session^4 -0.033907   0.043173  -0.785  0.43409
ConditionMassed:Session^5  0.031270   0.043173   0.724  0.47058
ConditionMassed:Session^6 -0.035551   0.043173  -0.823  0.41221
ConditionMassed:Session^7 -0.050430   0.043173  -1.168  0.24555
ConditionMassed:Session^8 -0.017639   0.043173  -0.409  0.68373
ConditionMassed:Session^9 -0.007453   0.043173  -0.173  0.86330
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07478 on 100 degrees of freedom
Multiple R-squared:  0.3673,    Adjusted R-squared:  0.2471
F-statistic: 3.055 on 19 and 100 DF,  p-value: 0.0001576
```