

An investigation of fingerprint submission  
decision making

Helen Christine Earwaker

UCL Centre for the Forensic Sciences

UCL Department of Security and Crime Science

UCL Department of Experimental Psychology

University College London

Submitted for the degree of Doctor of Philosophy in Forensic Science

Supervisors: Dr Ruth M. Morgan

Dr Adam J. L. Harris

Student thesis declaration:

I, Helen Christine Earwaker confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

Name:

Date:

## Acknowledgements

I would like to thank EPSRC (EP/G037264/1) for funding this research project. The funding was invaluable in enabling meetings, research sessions, and workshops with participating fingerprinting stakeholders across the UK, and for facilitating the presentation of the research findings at a number of key national and international meeting and conferences.

Thanks should also go to the UCL SECReT Doctoral Training Center within UCL Department of Security and Crime Science for giving me the opportunity to embark upon this project within such a richly diverse and multidisciplinary environment.

I would like to offer special thanks to my supervisors Dr Ruth Morgan (1<sup>st</sup> supervisor and Director of the UCL Centre for the Forensic Sciences) and Dr Adam Harris (2<sup>nd</sup> supervisor and Senior Lecturer in the UCL Department of Experimental Psychology) for their continued support and invaluable advice throughout the research process.

Special thanks should also go to Lisa Hall (Fingerprint Consultant within the Metropolitan Police Service) for crucial advice and encouragement at an early stage of the project, and also to Dr Gary Pugh for supporting participation in the research within the Metropolitan Police Service. I would also like to thank Nick Marsh (MPS) and Vaughn Sears (CAST) for their valuable time in assisting with the production of experimental materials. Indeed, I would like to thank all stakeholders and practitioner and examiner participants from all three participating police forces for their willingness and, indeed, enthusiasm to participate. Without their dedication and commitment to the project the present thesis would not have been possible.

I would like to thank all of the organisations who have invited me to talk about my research, including the Forensic Regulator's Fingerprint Quality Standards Working Group, and National Fingerprint Laboratory Group in the UK.

I would like to also offer my thanks to the Forensic Science Knowledge Transfer Network for allowing me the opportunity to carry out a piece of work that introduced me to a number of key stakeholders within UK fingerprinting and gave me a valuable wider view of the context of the present thesis.

Finally, I would like to thank my colleagues and friends within UCL Centre for the Forensic Sciences for providing an inspirational multidisciplinary learning environment and PhD support network.

## Abstract

This thesis investigates the submission of fingermark evidence from the fingerprint laboratory (where the enhancement of crime scene fingermarks occurs) to the fingerprint bureau (where examiners compare crime scene and suspect fingermarks) within UK forensic science. Initial research presented in this thesis identifies a discrepancy between laboratory practitioner mark submission decision making and the usability decisions made by fingerprint examiners, in the case of ambiguous fingermarks, leading to the potential for a loss of evidence that could be used to identify a suspect. Further empirical research explores the components of this decision process through consideration of decision success, cues, thresholds, and factors that influence the decision process. Qualitative research explores the rationale behind the mark submission decisions of practitioners, identifying a common reliance on a numerical value of characteristics present as a threshold for submission. The reliability of the use of a minutiae count as a method for increasing the objectivity of the submission is further investigated and variation between the minutiae cue detection of practitioners and examiners is identified. A contrast effect is found to occur in relation to practitioners making submission decisions concerning fingermarks in situ on exhibits that contain background marks, and this effect is discussed in relation to the differences in practitioner and examiner quality determination procedure. The findings of these empirical studies are presented and explained in terms of psychological theories of judgement and decision making, as well as in terms of their procedural and practical implications for fingerprint evidence recovery, and their wider implication within the holistic forensic process and criminal justice system.

## Table of Contents

Chapter 1	Introduction .....	16
1.1.	Context and importance of novel research .....	16
1.1.1.	Overall aims of thesis .....	17
1.2.	Thesis overview and research questions .....	17
Chapter 2	Literature Review .....	22
2.1.	The shifting identity of forensic science .....	22
2.2.	Introduction to UK fingerprinting: processes, stakeholders, and interactions .....	24
2.2.1.	Processes.....	24
2.2.2.	Stakeholders and interactions.....	25
2.3.	Communication.....	26
2.4.	The concept of individualization by friction ridge skin .....	29
2.5.	‘Cognitive forensics’ – a developing field .....	31
2.6.	Why investigate decision making within the practitioner fingerprint submission process?33	
2.6.1.	Potential differences in fingerprint sufficiency decision making between practitioners and examiners.....	34
2.6.2.	The importance of feedback .....	35
2.6.3.	Existing research into practitioner submission decision making .....	36
2.7.	The consideration of errors within forensic science .....	38
2.8.	The challenge of establishing ground truth.....	40
2.9.	The application of psychological theory to the fingerprint submission decision .....	41
2.9.1.	The potential biasing effect of crime context.....	41
2.9.2.	The relationship between confidence and accuracy in decision-making .....	43
2.9.3.	The application of signal detection theory to fingerprint sufficiency decision making	44
2.9.4.	The application of Brunswik’s Lens Model to the fingerprint submission process .....	49
2.9.5.	Visual perception within fingerprint quality assessment .....	50
2.9.6.	Effect of background context .....	51
2.10.	Summary of the objectives of this thesis .....	53

Chapter 3	An investigation of fingerprint submission decision making within UK police forces .....	57
3.1.	Introduction .....	57
3.1.1.	The potential for inefficiencies in the fingerprint laboratory mark submission process. .....	57
3.1.2.	Expanding the scope of previous research .....	58
3.2.	Summary of objectives.....	60
3.3.	Method .....	61
3.3.1.	Overview of method .....	61
3.3.2.	Method for objectives 3.1 – 3.3 .....	62
3.3.3.	Method for objective 3.4.....	70
3.4.	Results.....	70
3.4.1.	Data grouping.....	70
3.4.2.	Objective 3.1: Assessment of the efficiency of fingerprint submission decision-making within Laboratory A .....	71
3.4.3.	Objective 3.2: The effect of crime context on fingerprint submission decision-making within Laboratory A .....	78
3.4.4.	Objective 3.3: Self-reported confidence.....	84
3.4.5.	Objective 3.4: An assessment of inter-laboratory consistency in fingerprint submission decision-making.....	92
3.5.	Summary of key findings.....	101
3.6.	Discussion .....	103
3.6.1.	Objective 3.1: Assessment of the efficiency of fingerprint submission decision making within Laboratory A .....	103
3.6.2.	Objective 3.2: The effect of contextual information on practitioner submission decision making within Laboratory A.....	106
3.6.3.	Objective 3.3: Assessment of the relationship between self-reported confidence and decision accuracy.....	107
3.6.4.	Objective 3.4: A comparison of fingerprint laboratory fingerprint submission decision making between the Metropolitan Police Service and Laboratory A.....	109

3.7.	Conclusion.....	112
Chapter 4 An investigation of self-reported fingerprint quality assessment decision rationale.... 113		
4.1.	Introduction .....	113
4.2.	Summary of objectives.....	119
4.3.	Method .....	120
4.3.1.	Overview of method .....	120
4.3.2.	Participants .....	121
4.3.3.	Materials and procedure .....	121
4.3.4.	Analytical methodology.....	122
4.4.	Results.....	124
4.4.1.	Objective 4.1: Analysis of metacognition within Laboratory A.....	124
4.4.2.	Objective 4.2: Inter-laboratory consistency in decision rationale .....	136
4.4.3.	Objective 4.3: Fingerprint examiner usability decision rationale .....	146
4.5.	Summary of key findings.....	151
4.6.	Discussion .....	152
4.6.1.	Objectives 4.1 and 4.2: Practitioner fingerprint submission making self-reported decision rationale.....	152
4.6.2.	Objective 4.3: A comparison of practitioner submission decision rationale and examiner usability decision rationale .....	156
4.7.	Conclusion.....	160
Chapter 5 A comparison of examiner and practitioner minutiae counts, and an assessment of inter-practitioner variability..... 161		
5.1.	Introduction .....	161
5.1.1.	The importance of minutiae counts within the mark submission process .....	161
5.1.2.	The use of minutiae counts by fingerprint examiners .....	162
5.1.3.	Variation in performance in visual perception tasks .....	162
5.1.4.	The psychological basis for visual perception .....	163
5.1.5.	Potential differences between examiner and practitioner minutiae count .....	164

5.1.6.	The difference between a numerical standard for submission and a numerical standard for identification .....	165
5.2.	Summary of objectives.....	165
5.3.	Method .....	166
5.3.1.	Overview of method .....	166
5.3.2.	Materials .....	166
5.3.3.	Participants .....	170
5.3.4.	Procedure.....	170
5.4.	Results.....	172
5.4.1.	Aims of analysis.....	172
5.4.2.	Data preparation.....	172
5.4.3.	Objective 5.1: An examination of inter-practitioner variability in fingerprint minutiae count .....	172
5.4.4.	Objective 5.2: A comparison of variability between practitioners and examiners ....	177
5.4.5.	Objective 5.3: A comparison of practitioner and examiner minutiae counts .....	180
5.4.6.	Objective 5.4: The effect of fingerprint quality on minutiae count and variability ...	182
5.4.7.	Objective 5.5: Inter-laboratory consistency in practitioner variability and minutiae count .....	188
5.5.	Summary of key findings.....	189
5.6.	Discussion .....	191
5.6.1.	Variation in minutiae count.....	191
5.6.2.	Relationships between examiner and practitioner minutiae counts.....	192
5.6.3.	The effect of mark quality on variability and minutiae count .....	192
5.6.4.	Individual differences in visual perception.....	194
5.6.5.	Operational significance of these findings .....	196
5.7.	Conclusion.....	198
Chapter 6	The effect of background mark quality on mark submission decisions .....	199
6.1.	Introduction .....	199



6.1.1.	Differences in mark quality assessment practices between practitioners and examiners .....	199
6.1.2.	Contrast effects, assimilation effects, and the Inclusion/Exclusion Model .....	200
6.1.3.	Decision by sampling.....	202
6.2.	Summary of objectives.....	205
6.3.	Method .....	206
6.3.1.	Summary of methodological process and requirements.....	206
6.3.2.	Materials .....	206
6.3.3.	Participants .....	216
6.3.4.	Procedure.....	217
6.4.	Results.....	218
6.4.1.	Preparation and coding of data.....	218
6.4.2.	Summary of demographic information .....	219
6.4.3.	Objective 6.1. The main effect of context (background mark quality) .....	219
6.4.4.	Objective 6.2: Individual differences in mark submission .....	223
6.4.5.	Objective 6.3: Assessing for the influence of demographic factors and the presence of order effects .....	226
6.5.	Summary of key findings.....	227
6.6.	Discussion .....	228
6.6.1.	Objective 6.1: An investigation of the main effects of background mark quality on target mark submission .....	228
6.6.2.	Objective 6.2: Individual differences in mark submission .....	233
6.6.3.	Objective 6.3: Assessing for the influence of demographic factors and the presence of order effects .....	234
6.7.	Conclusion.....	235
Chapter 7	Discussion .....	236
7.1.	Overview of discussion.....	236
7.2.	Inefficiencies within the fingerprint submission process.....	236

7.3.	Using decision making models to explore why the fingerprint submission process is not efficient.....	237
7.3.1.	The principles of decision theory .....	237
7.3.2.	The components of signal detection theory.....	238
7.3.3.	Decision matching using Brunswik’s Lens Model .....	238
7.4.	Components and interactions within fingerprint submission decision making - a novel model .....	239
7.5.	Discussion of the key findings of this thesis in terms of the novel fingerprint submission decision model.....	242
7.5.1.	Decision Success.....	242
7.5.2.	Cues.....	246
7.5.3.	Thresholds.....	249
7.5.4.	Influences upon decision making .....	252
7.5.5.	Ecology.....	256
Chapter 8	Conclusion.....	261
8.1.	Overview of key findings .....	261
8.1.1.	Key findings of Chapter 3 .....	262
8.1.2.	Key findings of Chapter 4 .....	265
8.1.3.	Key findings of Chapter 5 .....	267
8.1.4.	Key findings of Chapter 6 .....	268
8.2.	Implications of the empirical findings of this thesis within fingerprint laboratory fingerprint submission decision making.....	269
8.3.	Implications of the findings of this thesis within the wider process of fingerprint evidence recovery .....	270
8.4.	Implications of the findings of this thesis for fingerprint and wider forensic science research .....	272
8.5.	Implications of the findings of this thesis for wider forensic science and the legal system ....	274
8.6.	Implications for other domains .....	275

8.7. Summary of research impact .....	276
List of publications .....	278
References .....	279
List of Appendices .....	294

## List of Figures

Figure 2.1. Fingerprint evidence processes, stakeholders, and influences (FoSci Community, 2014)	27
Figure 2.2. Sensitivity and specificity calculation for ROC analysis for fingerprint submission (adapted from Mickes et al, 2012)	47
Figure 2.3 - An example of a ROC curve	48
Figure 3.1. Example experimental fingermarks (not to scale)	63
Figure 3.2 - Formation of experimental image sets	64
Figure 3.3 - An overview of the experimental method	67
Figure 3.4 - Experimental methodology	69
Figure 3.5 - Pie chart of overall decision outcomes within Laboratory A	72
Figure 3.6 - Decision outcomes according to examiner mark assessment	73
Figure 3.7 - Decision outcomes according to experimental fingermark	75
Figure 3.8 - Decision outcomes according to practitioner	76
Figure 3.9 - Pie chart illustrating decision consistency	80
Figure 3.10 - Decision consistency according to experimental fingermark	82
Figure 3.11 - Decision consistency according to practitioner	83
Figure 3.12 - A comparison of self-reported confidence according to decision category	86
Figure 3.13 - Boxplot illustrating self-reported confidence according to experimental fingermark	88
Figure 3.14 - Boxplot illustrating self-reported confidence according to practitioner	89
Figure 3.15 - Scatterplot of mean accuracy against mean decision confidence for decisions made in relation to each experimental fingermark	90
Figure 3.16 - Scatterplot of mean accuracy against mean decision confidence according to practitioner	91
Figure 3.17 - Comparison of overall decision outcomes according to laboratory	93
Figure 3.18 - Comparison of practitioner decision outcome profiles of MPS and Lab A	95
Figure 3.19 - Comparison of decision outcomes according to laboratory and crime context	96
Figure 3.20- Mean decision accuracy and mean decision confidence for each practitioner from both laboratories	99
Figure 3.21 - Comparison of collaboration per experimental fingermark	100
Figure 4.1 - A fingerprint practitioner decision to 'submit' described in the language of signal detection theory	113
Figure 4.2 - An adaptation of Brunswik's Lens Model to illustrate the matching of the quality cue weights - Adapted from Newell and Shanks (2014)	116

Figure 4.3 - Word cloud illustrating all decision factors provided by Laboratory A practitioners, and the relative frequency of these factors -----	126
Figure 4.4 - Bar chart illustrating the total occurrence of decision factors within Laboratory A -----	132
Figure 4.5 - The composition of decision factors within the 2nd level detail category -----	133
Figure 4.6 - Incidences of decision categories according to Laboratory A practitioner-----	134
Figure 4.7 - Make up of decision factors according to experimental fingerprint -----	135
Figure 4.8 - 'Word cloud' displaying MPS decision factors and their relative frequency within the data -----	138
Figure 4.9 - Occurrence of decision categories within the MPS Laboratory-----	142
Figure 4.10 - A comparison of the make-up of decision categories between Laboratory A and MPS Laboratory decisions -----	144
Figure 4.11 - A comparison of decision categories according to laboratory and crime context-----	145
Figure 4.12 - Occurrence of decision factors within each decision category for examiner usability decisions -----	149
Figure 4.13 - A comparison of the make-up of decision rationale categories between practitioners and examiners-----	150
Figure 4.14 - A novel adaptation of Brunswik's Lens Model including key decision cues gleaned from empirical study -----	159
Figure 5.1 - Example of experimental fingerprints of each assigned quality ranking (enlarged image, not to scale) -----	168
Figure 5.2 - Frequency of ranges of practitioner minutiae count-----	175
Figure 5.3 - Box plots of practitioner (Figure 5.3. A) and examiner (Figure 5.3. B) minutiae counts for all fingerprints-----	176
Figure 5.4 - Total number of minutiae observed by each participant in all experimental tasks-----	177
Figure 5.5 - A comparison of practitioner and examiner standard deviation in minutiae count -----	179
Figure 5.6 - A comparison of practitioner and examiner mean minutiae count per fingerprint -----	183
Figure 5.7 - Mean total of minutiae according to mark quality and participant job role-----	184
Figure 6.1- Summary of experimental methodology -----	207
Figure 6.2- Example of a labelled photograph file containing a mixture of good quality and poor quality marks -----	209
Figure 6.3- The background image used to create all mock exhibits-----	210
Figure 6.4 - The content of mock exhibits-----	211
Figure 6.5 - An example of a mock exhibit -----	213
Figure 6.6 - Example of a marked up exhibit -----	218

Figure 6.7 - Box and whisker plots illustrating spread of number of target marks submitted according to proportionate mark quality ----- 221

Figure 6.8 - Relationship between target and background mark submission----- 223

Figure 6.9 - Total number of marks submitted by each practitioner according to context mark quality ----- 224

Figure 6.10 - Total number of background marks submitted per participant----- 225

Figure 7.1 - Components and interactions within fingerprint submission decision making - a novel model----- 241

Figure 8.1 - Relationship between thesis aims and key findings----- 262

## List of Tables

Table 2.1 - Summary of thesis key objectives and current knowledge gaps -----	56
Table 3.1 - Justification of aspects of material production-----	65
Table 3.2 - Contingency table for Chi Squared test for difference in proportion of erroneous and agreement outcomes for sufficient and insufficient fingermarks -----	72
Table 3.3 - Tabulated error rates, sensitivity, and specificity -----	74
Table 3.4 - Tabulated false positive and false negative discovery rates-----	74
Table 3.5 - Descriptive statistics in relation to decisions made by all practitioners -----	76
Table 3.6 - Contingency table for Chi Square test for difference in expected and observed decision outcomes according to time of presentation-----	78
Table 3.7 - Contingency table for Chi Square test for relationship between crime context and decision outcome-----	79
Table 3.8 - Descriptive statistics in relation to self-reported confidence -----	85
Table 3.9 - Tabulated results of Wilcoxon signed ranks tests comparing self-reported confidence---	87
Table 3.10 - Contingency table for Chi Square test of relationship between laboratory and decision outcome-----	93
Table 3.11 - Comparison of descriptive statistics in relation to all decisions made by MPS and Laboratory A practitioners-----	94
Table 3.12 - Tabulated comparison of decision consistency between MPS and Lab A -----	97
Table 3.13 - Contingency table for Chi Square test for difference in type of decision inconsistency according to laboratory-----	98
Table 4.1- The categorisation and occurrence of Laboratory A decision factors-----	128
Table 4.2 - Decision factors making up decision categories (including frequencies) for MPS data --	139
Table 4.3 - Decision factors and categories for examiner usability determinations-----	147
Table 5.1 - Order of experimental image sets -----	169
Table 5.2 - Descriptive statistics for practitioner minutiae count -----	174
Table 5.3 - Descriptive statistics for examiner minutiae count-----	178
Table 5.4 - Mean minutiae count (per fingerprint) values for practitioners and examiners -----	181
Table 5.5 - Mean minutiae count (per participant) values for practitioners and examiners -----	181
Table 5.6 – Descriptive statistics for experimental marks ranked according to mean practitioner minutiae count -----	185
Table 5.7 - Descriptive statistics for experimental marks ranked according to mean examiner minutiae count-----	187
Table 5.8 - Descriptive statistics for practitioner minutiae count, according to laboratory -----	189

Table 6.1 - The application of decision by sampling theory to fingerprint submission (Adapted from Stewart et al. 2009)-----	204
Table 6.2 - Key to counterbalancing groups for target mark positioning -----	214
Table 6.3 - Order of quality series for each category group-----	214
Table 6.4 - Key to make up of quality series -----	214
Table 6.5 - Counterbalancing conditions assigned to each participant-----	215
Table 6.6 - Descriptive statistics for numbers of target marks submitted for each laboratory -----	220
Table 6.7 - Pairwise comparisons testing for significant differences-----	222
Table 6.8 - Tabulated results of mixed between-within ANOVA analyses testing for interaction and between subject effects according to key demographic factors-----	227



# Chapter 1 Introduction

## 1.1. Context and importance of novel research

Fingerprint laboratory practitioners carry out the vital role of visualising latent fingermarks on exhibits at, and recovered from, the crime scene using a range of chemical and physical development techniques. Once visualised, it is the responsibility of the laboratory practitioner to determine whether any fingerprint ridge detail present is of sufficient quality to be submitted to a fingerprint examiner within a Fingerprint Bureau, for search or comparison against a fingerprint from a person of interest in a legal case. This fingerprint submission decision currently lacks an objective methodology or guidelines, and as such can be considered to be a subjective quality judgement made by a human decision maker.

‘Human factors’ have been frequently discussed in relation to fingerprint evidence in recent years. Such discussion and empirical study, however, has focussed upon the aspects of fingerprint evidence process traditionally considered to be interpretative in nature. This has primarily involved the Analysis, Comparison, Evaluation, and Verification processes (commonly referred to as ACE-V) carried out by fingerprint examiners. Empirical study within these areas has found examiners carrying out the ACE-V process to be vulnerable to a number of subconscious psychological effects, commonly referred to in the literature as ‘cognitive biases’, and a number of potential solutions have been proposed to increase the reliability of aspects of the ACE-V process carried out by fingerprint examiners.

Empirical study in relation to the fingerprint submission process, which, arguably is similar to the Analysis stage of the examiner ACE-V process, has, however, been largely overlooked. It is crucial that the subjective decisions made within the fingerprint laboratory are considered as an important aspect of the fingerprint recovery process as faulty decision making at this point could mean the loss of crucial evidence at this early stage which could have led to an identification, or could mean wasted resources within a case.

The role of the laboratory practitioner has also been traditionally viewed as a more technical and less interpretative than that of the fingerprint examiner. As such the focus of the training provided for practitioners at a national level has focussed upon the selection and application of the most appropriate fingerprint development techniques, and consideration towards training and expertise development in mark ‘Analysis’ have been largely neglected.

It is essential to consider this decision making occurring early in the progression of fingerprint evidence in order to ensure the efficiency and transparency of the entire evidence recovery process particularly

given the lack of training leading to expertise in this key task, and the lack of empirical study investigating this process.

#### 1.1.1. Overall aims of thesis

This thesis aims, therefore, to investigate decision making within the fingerprint development laboratory, in particular, in relation to the decision to submit or discard items of developed fingerprint ridge detail, as per the objectives outlined in section 1.2.

### 1.2. Thesis overview and research questions

The present thesis takes the following approach to investigating decision making within the fingerprint laboratory:

Chapter 2: Literature review.

A review of the relevant literature begins by describing the cultural shift towards a more scientific culture within the forensic sciences. The processes, stakeholders, and interactions within UK fingerprinting are introduced. The concept of 'cognitive forensics' and the study of decision making within forensic science and fingerprinting are discussed. The key decision process of fingerprint submission from the fingerprint visualisation laboratory to the fingerprint bureau to be empirically studied within this thesis is introduced and psychological theories of judgment and decision making pertinent to the fingerprint submission decision process are summarised.

Following the overarching review of the relevant literature, four independent chapters (Chapters 3 – 6) will each seek to answer a distinct research question and target a separate aspect of the fingerprint submission decision. Each of these four chapters will begin with a specific introduction including a review of literature pertinent to that area, a summary of the research questions to be investigated, and will include the method, results and discussion of an empirical study in the area described.

Chapter 3: 'An investigation of fingerprint submission decision making within UK police forces'

Chapter 3 introduces key previous research by Earwaker et al. (2015) that identified inefficiencies within the process of fingerprint submission from the fingerprint visualisation laboratory to the fingerprint bureau of the Metropolitan Police Service (MPS), in terms of the usability of the fingerprints submitted and discarded by laboratory practitioners. The inappropriate determination of the usability of fingerprints suggested the possibility of identifiable fingerprints being lost from the evidential chain. Differences in structure and training between the MPS and other UK police forces are described and the resultant need to establish the reproducibility of the current research within

other UK police forces is set out. Empirical study contained within this chapter aims to meet the following objectives:

Objective 3.1 – Assessment of the efficiency of fingerprint submission decision making within Laboratory A

To establish the extent to which fingerprint laboratory fingerprint submission decisions match the usability determinations of fingerprint examiners within a non-MPS UK police force (Laboratory A: anonymised for inclusion within this thesis).

Objective 3.2 – Assessment of the effect of contextual information on submission decision making within Laboratory A

To establish the effect of case context (crime type) on fingerprint submission threshold.

Objective 3.3 – Assessment of the relationship between self-reported confidence levels and decision accuracy within Laboratory A

To establish the relationship between the self-reported confidence and accuracy of practitioner fingerprint submission decisions.

Objective 3.4 – An assessment of inter-laboratory consistency in mark submission decision making

A comparison of fingerprint submission decision making between Laboratory A and that of the Metropolitan Police Service Serious Crime Fingerprint Laboratory (as documented in Earwaker et al, 2015), in relation to mark submission efficiency, the effect of case context, and the relationship between self-reported confidence and decision accuracy.

The empirical study described within this chapter establishes similarities in the overall inefficiencies within fingerprint submission decision making between the two laboratories and within the relationship between confidence and accuracy, however, establishes a difference in the effect of case context on fingerprint submission decision making. These findings are discussed in terms of signal detection theory and the cognitive forensic literature.

Chapter 4: An investigation of self-reported fingerprint quality assessment decision rationale  
Building upon the findings of potential inefficiencies in the mark submission process within UK police forces, established within Chapter 3, Chapter 4 further explores the process of practitioner and examiner mark quality assessment decision making. The respective submission and usability determinations of practitioners and examiners are described according to a novel adaptation of Brunswik's Lens Model, highlighting the importance of decision cues within the mark submission

process. A primarily qualitative empirical study is carried out through the use of a grounded theory style approach to data analysis, in order to meet the following objectives:

Objective 4.1. To explore the rationale behind laboratory practitioner fingerprint submission decision making within Laboratory A.

To explore the self-reported reasons for laboratory practitioner fingerprint submission decisions including the assessment of inter-practitioner and inter-fingerprint variation in fingerprint submission decision rationale.

Objective 4.2. An investigation of inter-laboratory consistency in practitioner decision rationale  
To establish the decision rationale within the MPS laboratory and compare this to that of Laboratory A. The effect of crime context on decision rationale is also compared between the two laboratories, as differences in the effect of crime type of mark submission decision outcomes were established within Chapter 3.

Objective 4.3. Exploring the relationship between practitioner submission decision rationale and fingerprint examiner usability rationale

To establish fingerprint examiner usability decision rationale and to compare these rationale with those established in relation to the fingerprint submission decision of laboratory practitioners.

The results of the primarily qualitative analysis presented within this chapter highlight the importance of second level detail, and, more specifically minutiae count (the number of characteristics deemed to be present within a fingerprint) within both practitioner submission decision making and examiner usability decision making.

Chapter 5 – ‘A comparison of examiner and practitioner minutiae counts and inter-practitioner variability’

Chapter 5 builds upon the findings of Chapter 4 by focussing upon the use of minutiae present within a fingerprint as a mechanism for submission decision making. Chapter 5 provides a novel investigation of inter-variability in minutiae count within laboratory practitioners, and examines the relationship between practitioner and examiner minutiae count. The following key objectives are addressed:

Objective 5.1 – an examination of inter-variability in fingerprint laboratory practitioner minutiae count

To establish the level of inter-variability in minutiae count between fingerprint laboratory practitioners.

Objective 5.2 – a comparison of inter-variability of practitioners and examiners

To compare the level of inter-variability in minutiae count between fingerprint laboratory practitioners and fingerprint examiners.

Objective 5.3 – a comparison of practitioner and examiner minutiae counts

To compare laboratory practitioner and examiner minutiae counts in relation to the same fingerprints within varying quality categories.

Objective 5.4 – a comparison of inter-variability according to fingerprint quality

To compare examiner and practitioner inter-variability in minutiae count according to the quality of fingerprints presented.

Objective 5.5 – an assessment of inter-laboratory differences in minutiae count and inter-practitioner variability

To compare minutiae, count and inter-variability between two different UK fingerprint recovery laboratories.

The empirical study reported within this Chapter establishes a high level of inter-variation in minutiae count within both practitioners and examiners and finds no significant difference in the overall minutiae counts of the two groups. This suggests that a strong reliance on using a numerical threshold for practitioner fingerprint submission may be problematic.

Chapter 6 – ‘The effect of background mark quality on fingerprint submission decisions’

Chapter 6 introduces the procedural differences in the laboratory practitioner fingerprint submission decision and the fingerprint examiner quality usability decision. The laboratory practitioner often makes their submission decision in relation to a ‘target’ fingerprint in situ upon an exhibit containing other background marks, whilst an examiner will often view an image of the submitted mark in isolation. The potential of the quality of the background marks upon an exhibit to affect the quality assessment of a target mark is discussed in terms of the theories of Decision by Sampling and Contrast and Assimilation Effects. Novel empirical study investigates the effect of background mark quality in relation to target mark submission in order to meet the following objectives:

Objective 6.1. - An investigation of the main effects of background mark quality on target mark submission

To establish the effect of background mark quality of the submission decisions made by laboratory practitioners in relation to ambiguous ‘target’ marks.

Objective 6.2. An assessment of individual differences in mark context effects

To investigate individual differences in the effect of the quality of background marks.

Objective 6.3. Assessing for the influence of demographic factors and the presence of order effects

To assess for any relationship between the effect of background mark quality and key demographic and experimental factors.

The main results of the empirical study outlined within Chapter 6 suggest that a contrast effect occurs in the case of extreme background quality exemplars with practitioners submitting more ambiguous target marks when background marks were of a high quality, and submitting less target marks when background marks were of a low quality

Chapter 7 provides a discussion of the key empirical findings of the thesis. A novel model for the consideration of the practitioner submission decision is introduced. This model combines the key elements of the practitioner fingerprint submission decision process defined as what is meant by decision success, and how cues, thresholds and influences interact within the decision making process. The key findings of Chapters 3-6 are all discussed in terms of the key elements of this novel model. The model draws upon key psychological principles but remains accessible to a fingerprint practitioner audience.

Chapter 8 provides a conclusion to the thesis , summarising the key findings of the empirical studies within the thesis (reproducible inefficiencies in the fingerprint submission process, the varying effects of crime context on submission decision making, a self-reported reliance on minutiae counts and thresholds when making a submission decision, a large level of inter-practitioner variation in minutiae count and the lack of a clear relationship between practitioner and examiner minutiae count, and a contrast effect according to extreme background mark quality within an exhibit). The conclusion explores the impact and potential operational significance of the findings within UK fingerprint submission, additional areas of forensic science practice, and the transferable application of these findings within other domains.

# Chapter 2 Literature Review

## 2.1. The shifting identity of forensic science

In 1993 the validity of the judicial end product of criminalistics was called into question, based upon its theoretical scientific foundations. As a result of *Daubert v Merrell Dow Pharmaceuticals* (1993) the US Supreme Court introduced a new admissibility standard for scientific evidence, stating that scientific testimony must be based upon defensible scientific foundations (Saks, 2010). Judges in some US states must now consider whether, and how, scientific principles have been tested, and what the results were (Saks, 2010). This Daubert admissibility criteria and the UK report by The Law Commission (2011) have subjected forensic science to a 'first principles scientific scrutiny' (Saks, 2010), essential to it fulfilling its role within the legal system.

Defending such 'first principles' has been exposed as problematic for forensic science which has come to rely on unquestioned legal acceptance (Cole, 2011). The two founding principles of forensic science are well established as being the 'principle of exchange' by Locard (Crispino et al., 2011, Ashbaugh, 1999), and 'concept of individuality' by Kirk (Crispino et al., 2011). These are bold claims which are difficult, if not impossible to empirically establish, leading to the suggestion that forensic science lacks a scientific basis (Saks and Koehler, 2007), although both claims are, arguably, philosophically admissible as science as they are falsifiable in principle (Crispino et al., 2011).

Forensic scientists must also validate the methodology used to apply these founding principles for the means of individualizing. From its founding principles forensic science has evolved into the practical application of criminalistics; 'the science of individualization' (Kaye, 2003). This evolution has, arguably, blurred the necessary distinction between the symbiotic disciplines of forensic science and criminalistics. Saks and Koehler (2008) refer to the necessary paradigm shift of forensic science from a pre-science to an empirically grounded science, whilst highlighting the differences between 'traditional' scientists who stem from an epistemological background, and forensic scientists who often have no prior scientific background aspire to less stringent set of scientific norms, whilst Houck (2013) comments on the restrictions imposed by the conflicting budgetary constraints of the law enforcement agencies which employ them. It seems that whilst the majority of forensic practitioners are not traditional scientists, but are experienced in the practical application of methodology to casework, there is a natural division here, with the forensic scientist primarily scientifically trained responsible for the validation of theories, principles and techniques, leaving the practitioner (or criminalist) to ply their expertise in the practical application of this forensic science for the purposes

of individualisation, until such a time that budgetary constraints may allow the two skill sets to overlap in the workplace.

Until forensic science has made the epistemological steps to validation of the field, criminalists need to be more cautious of unjustifiable claims that threaten the integrity of the discipline and its ability to provide a valuable legal contribution. Being transparent in this way has been shown not to damage the impact of the evidence given (Chisum and Turvey, 2000), but ensures admissibility of evidence, without which forensic science is arguably futile.

Steps have recently been taken to improve the level of external innovation and academic input to the forensic sciences in an attempt to begin to bridge the gap between science and forensic practice. In the UK the need for academic and practitioner collaboration has been acknowledged through the publication of the Silverman review (B. Silverman, 2011), and a report by the House of Commons Science and Technology Committee (2013) which suggests a growing momentum to look for a collaborative and research focussed 'path forward' in the UK to tackle the 'paucity of research in forensic science' highlighted by the NAS (National Research Council, 2009 p186). Equally, recent publications by the UK Forensic Science Regulator (Tully, 2015), who is responsible for ensuring that quality standards are maintained across UK forensic science, stress the need for validation and a collaborative approach.

Indeed, there is a drive within forensic science to ensure that academic research is fit for purpose and meets the needs of a law enforcement client. As such there has been an increase in the level of communication and collaboration between industry and academia. Such communication was invigorated by the LiveTime Forensics document published by the Association of Chief Police Officers (ACPO) (Association of Chief Police Officers, 2012). This document specified the areas in which operational policing and forensic science would benefit from external innovation in order to target transferable knowledge and academic research to meet business needs.

In 2012 the Forensic Science Knowledge Transfer Network (originally named the Forensic Science Special Interest Group) was established in response to recommendation made in the Silverman review (B. Silverman, 2011). The community, which is backed by the Government Technology Strategy Board and is run by the Knowledge Transfer Network, encompasses all stakeholders involved in forensic science, including end users, suppliers of products or services, academics, and policy makers (Forensic Science Special Interest Group, 2014) with the primary aim of "enabling closer networking and better communication between all forensic science stakeholders for improved research and development" (Forensic Science Special Interest Group, *ibid*). The steering group of the Forensic Science Community is made up of operational, policy maker, and academic stakeholders from across the forensic domain



to ensure a balanced approach, and is attended by the Forensic Science Regulator. One of the key outputs of the community to date has been the Challenges Catalogue. Published on the website of the Forensic Science Knowledge Transfer Network, the catalogue allows operational stakeholders to detail the current challenges that they face which they feel would benefit from the application of external innovation or academic research. In turn, researchers and innovators can scan for capability gaps and collaborative opportunities (Forensic Science Special Interest Group, *ibid*). The Policing and Crime Reduction Research Map provided on the College of Policing (CoP) 'What Works' website also provides a database of forensic science research which can help to provide operational stakeholders with up to date information in relation to current avenues of relevant academic research and can also facilitate collaborative research projects across academic institutions.

In addition, there has been a recent move towards the establishment of partnerships between UK university forensic science departments and UK police forces. A good example of such a relationship is the Forensic Innovation Centre which is the result of a collaboration between the Institute of Criminal Justice Studies at the University of Portsmouth and Hampshire Police Force and, more recently, Hampshire Fire Service (Forensic Science Special Interest Group, 2014). Such partnership allows the sharing of resources in terms of laboratory facilities, the development of targeted research projects of direct benefit to operational needs, and placement opportunities and real world experience for students. This type of collaboration allows the true integration of science and practice.

It would appear that forensic science practitioners in the UK are beginning to move away from a silo existence and become more transparent. An increase in collaboration and knowledge transfer and an openness to external innovation will pave the way for the scientific and multidisciplinary input that is necessary to maintain the integrity of this societally essential discipline.

## 2.2. Introduction to UK fingerprinting: processes, stakeholders, and interactions

Fingerprints have been utilised as a trusted method of identification within a forensic context for over a hundred years (Champod et al. 2004). Today fingerprinting is a rapid and cost effective means of identification used widely within criminal cases in the UK.

### 2.2.1. Processes

Fingerprinting for forensic purposes can primarily be divided in to two key functional entities; fingermark visualisation, and fingerprint comparison. Together these two disciplines facilitate the progression of fingerprint evidence from crime scene to court.

Fingermark visualisation is carried out within the fingermark visualisation laboratory (also known as the Fingerprint Development Laboratory). Items of evidence from scenes of crime are submitted by scene of crime officers to the laboratory. The most appropriate chemical and physical development techniques are selected and carried out for each item of evidence, depending upon the type of surface present in accordance with recommended best practice provided by the Home Office Centre for Applied Science and Technology (CAST) (Bowman, 2014). Laboratory practitioners must then visually record (commonly by photography) fingermarks that they regard to be of sufficient quality for submission to a fingerprint examiner for comparison against a person of interest in the case. Fingerprint laboratory practitioners in the UK will either have been trained in house by their employing police force or through completion of the College of Policing Fingerprint Laboratory Officer training programme. Both training methods focus upon the selection and practical application of techniques to visualise latent fingermarks and their subsequent recovery (Lagden, 2014). It is a requirement that all fingerprint recovery laboratories must be accredited by the United Kingdom Accreditation Service (UKAS), and laboratories must pass annual inspections to remain accredited and thus operational.

The process of fingerprint comparison (more accurately described as ACE-V; Analysis, Comparison, Evaluation and Verification (SWGFAST, 2011)) is carried out by fingerprint examiners (also known as fingerprint experts) working within a Fingerprint Bureau. Examiners are responsible for comparing crime scene fingermarks (fingermarks recovered from crime scene or those visualised in the laboratory) with exemplar prints from persons of interest in a case. This process involves 'analysing' the crime scene fingermark by locating and annotating the minutiae present, and determining whether there is sufficient information present within the mark to proceed to carry out a comparison. The annotated fingermark is then 'compared' to the minutiae present in the exemplar print, the extent of the similarities is 'evaluated' and a decision is reached as to whether there is sufficient information present to suggest a match between the mark and the print. This decision is then corroborated by an additional examiner during the 'verification' process (SWGFAST, 2011).

It may be the case that fingerprint examiners are asked to present their conclusions in court should their findings be pertinent to an investigation, and as such, are expected to act as expert witnesses within their field. In addition, laboratory practitioners are also vulnerable to the scrutiny of the court and may be required to justify and explain the processes that they have carried out.

### 2.2.2. Stakeholders and interactions

There are a number of key stakeholders who have an influence on the fingerprint evidential process. The key processes, stakeholders, and interactions involved in the progression of fingerprint evidence from crime scene to court are represented in Figure 2.1. This diagram was developed on behalf of the

UK Forensic Science Knowledge Transfer Network (previously known as the Forensic Science Special Interest Group) (Forensic Science Special Interest Group, 2014) with the purpose of informing an audience of potential external innovators. The key stakeholders influencing the fingerprint evidential chain have been classified as relating to either; policing and law enforcement, regulation, accreditation, and training, research and development, or, industry. Additional job roles within fingerprint enhancement, recovery, and comparison are outlined in Appendices A.1, A.2, and A.3 respectively.

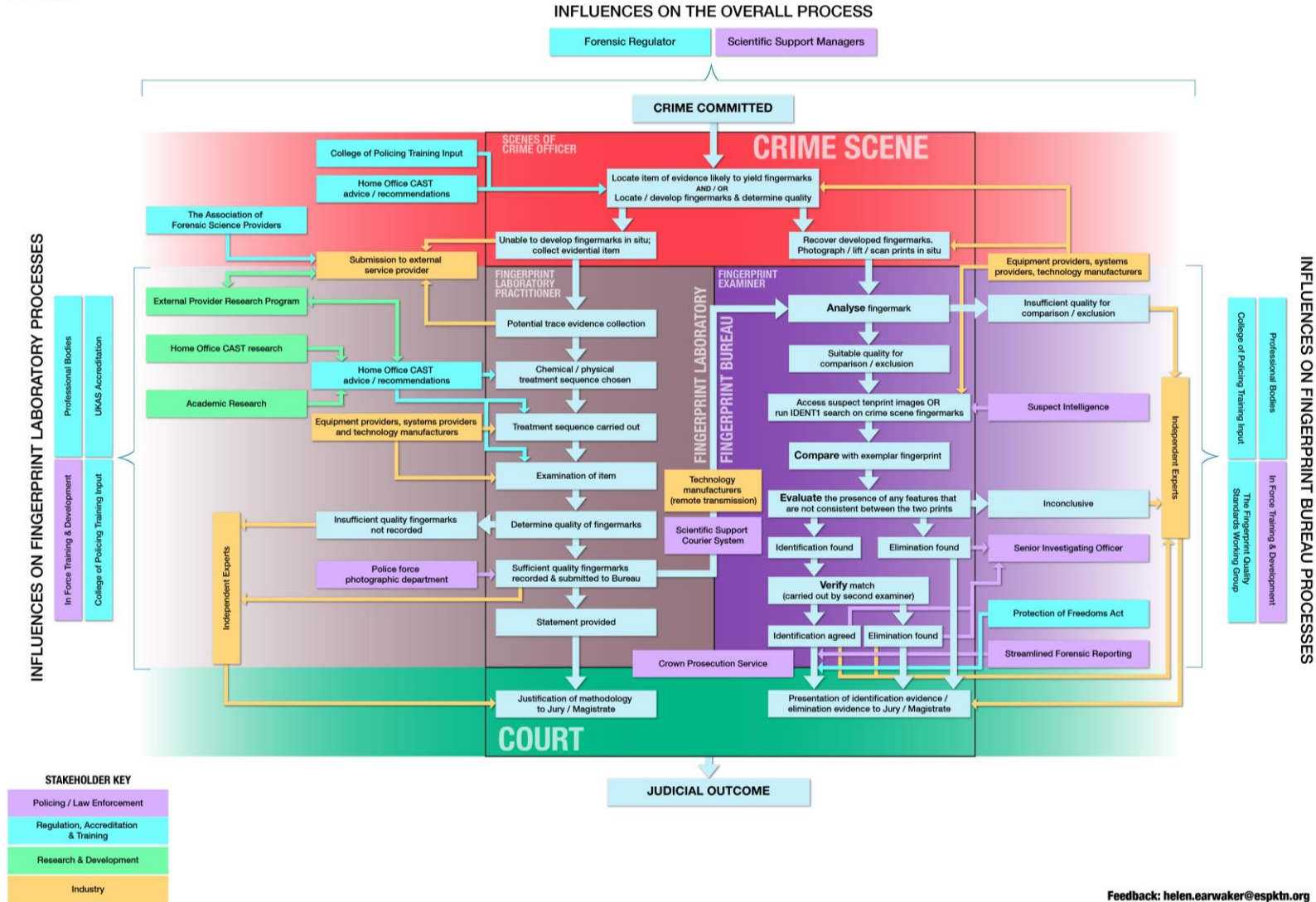
### 2.3. Communication

Within the UK it is routinely the case that both the laboratory and the bureau are situated within specialist scientific support departments within each UK police force, or in some cases, as a shared resource between two or three police forces. The lab and bureau within each force, or collaboration of forces, will interact differently according to a number of factors. Some laboratories and bureaux are geographically isolated from one another limiting the possibility of interaction between staff members within each unit. In other forces the laboratory and bureau may be situated within the same site or building increasing the opportunities for communication and collaboration. Policy and working practices will also play a role in the working relationship; in some police forces the relationship between the laboratory and the bureau may be more hands-on with examiners second checking marks allocated by the laboratory for submission, or even selecting marks for comparison themselves, whereas in other forces the workflow of the two units is entirely separate and the bureau acts only as the customer of the laboratory. Politics and sensitivities surrounding the job roles of examiners and practitioners have played a part in a typically limited working relationship. Historically it was the case that all fingerprint work was carried out within one unit, where practitioners were considered to carry out a basic, mechanical role of chemical treatment, with the expertise lying foremost with the examiners. Practitioners fought for fingerprint development to be considered as a specialism in its own right, leading to the creation of the laboratory as a separate, distinct entity. Whilst this ensured that laboratory practitioners were considered to be skilled in their own right it led to a divide in workflow and, resultantly, communication between the two specialisms (Stow, 2014).

Unlike some other jurisdictions, fingerprint work in the United Kingdom is undertaken at a local rather than national level. Each police force is responsible for setting its own policies, procedures and workflow in relation to the chemical development, submission and comparison of crime scene marks. Equally, each force determines the type of training provided to their fingerprint staff and prescribes the extent to which the lab and bureau work collaboratively. This local approach to working has

Figure 2.1. Fingerprint evidence processes, stakeholders, and influences (FoSci Community, 2014)

**FINGERPRINT EVIDENCE PROCESSES, STAKEHOLDERS AND INFLUENCES**  
JANUARY 2014



resulted in considerable variation in working practices and buy in of technology and systems (Charlton, 2014). With law enforcement agencies working within increasingly tight budgets, policy that directs resources to where they can have maximum evidential impact is increasingly important, and it is often the responsibility of local police forces or fingerprint units to make these policy decisions in relation to fingerprint work both in the UK and abroad. Guidelines in the US, for example, state that each forensic organisation should have its own policy on what makes a fingermark 'sufficient' for comparison during the ACE-V process (SWGFAST, 2011). Similar guidelines were also recently issued in relation to European laboratories (ENSFI, 2015). From a UK perspective Charlton (2013) highlights the importance of working towards standards of expediency which provide a balance between what is scientifically desirable and what is affordable. The Codes of Practice and Conduct issued by the UK Forensic Science Regulator also aims to ensure consistent quality standards across UK fingerprinting but still refers to the use of documented local policy and procedure (Forensic Science Regulator, 2016). The local differences in fingerprint service management in these areas suggests that there is potential for local variation in output and performance.

National fora for communication do exist within UK fingerprinting but there can be seen to have been failings in the inclusive nature of these as well as in the dissemination of information from them within recent years. There are two groups that sit at a national level and have an influence upon the output of forensic fingerprinting; The Fingerprint Quality Standards Working Group and the Fingerprint Strategic Network. The Fingerprint Quality Standards Working Group was established as a result of recommendations made following the Scottish Fingerprint Enquiry with the primary purpose of producing a fingerprinting annex to the Forensic Science Regulators codes of practice and conduct in order to ensure the reliability of fingerprint evidence (Forensic Science Regulator, 2014). The working group consists of a number of key stakeholders including the Forensic Science Regulator, Heads of a number of Fingerprint Bureaux and a representative from academia and UKAS. As such, the focus of the groups is quality standards within fingerprint comparison, without input from a fingerprint development stance. The Fingerprint Strategic Network is a national forum for communication for fingerprint bureaux, attended by Heads of Bureaux (Hall, 2014). Until recently there was currently organised national or local forum for fingerprint development laboratories (Bleay, 2014). The lack of a dedicated channel of communication for laboratories and a lack of representation on the national fingerprint groups may have presented a barrier to a consistent approach to working across laboratories and to the sharing of knowledge between laboratories and bureaux. It would seem to be of paramount importance that changes in quality standards discussed in relation to fingerprint comparison are communicated to laboratories as it is primarily the laboratory that is responsible for supplying the bureau with the raw material. If the requirements of the bureau changed due to new

ways of working designed to increase quality, then it would seem beneficial for the laboratory to be aware of these novel requirements so as to ensure that they are being met. For example, a change in bureau policy to statistically report similarities between fingerprints as opposed to simply reporting an identification or exclusion may have an impact on the quality of fingermarks that could be considered useable and should be submitted by the laboratory. A history of a lack of communication between laboratories at a national level and the absence of centrally agreed policy would suggest considerable differences in procedure, workflow, and systems between fingerprint units across the country, and contributes to the potential for local variation in efficiency and performance.

#### 2.4. The concept of individualization by friction ridge skin

The ability of fingerprints to theoretically individualise is justified based upon the uniqueness and permanence of friction ridge skin (Champod et al, 2004), resulting from multiple factors involved in the embryological formation of epidermal ridges (Ashbaugh, 1999, Kaye, 2003). Identification is based upon the assumption that the patterns formed by the epidermal ridges on every individual finger are unique and that they are distinguishable from the patterns on every other finger (Haber and Haber, 2008). However, there is a lack of empirical evidence to confirm the statistical nature of this uniqueness (Kaye, 2003). Empirical evidence for the statistical uniqueness of fingerprints, which would remove current inductive reasoning, could only be achieved by comparing the friction ridge skin of everyone who had ever lived; an unattainable goal. An attempt made to provide a statistical value for uniqueness for litigation purposes remains unpublished in the scientific press and has been criticised by Kaye (2003) for being inaccurate and badly designed. Even if statistical validation of theoretical uniqueness was achievable the practical ability to individualise from a fingerprint would still be dependent upon the methodology used.

Fingerprint comparison methodology through the implementation of the ACE-V process (SWGFAST, 2011) is a subjective and opinion driven discipline. This process has not been scientifically verified (Haber and Haber, 2008) and a lack of evidence that fingerprint examiners do indeed have expertise in identifying the source of fingermarks has been highlighted by Vokey et al. (2009). However, fingerprint practitioners and the courts still uphold a 'strong faith in uniqueness and (virtual) freedom from error' (Saks, 2010, p14), continuing to claim to individualise (Cole, 2014).

Error has, however, been shown to occur, most publicly with the misidentification of the Madrid bomber, but also in a number of other cases of misattribution (Cole, 2005). DNA analysis techniques, which have achieved statistical validity for identification, can now exceed the reliability of fingerprint evidence, leading to an increased overturning of fingerprint convictions, as in the case of Stephen Cowens (Cole, 2005, 2006). This illustrates the paradoxical position that 'good' evidence holds; with

less other evidence able to challenge it, it is less challenged and so strengthened, meaning that flaws can go undetected (Cole, 2006). This may suggest the potential for further cases of wrongful convictions based upon erroneous fingerprint evidence. The terminology and meaning of error within forensic science is discussed further within section 2.7.

There is now healthy, multidisciplinary debate about how to ensure the continued legal acceptance of fingerprint evidence and increase its validity (Koehler, 2008), with approaches divided between the validation of ACE-V methodology or basing testimony solely on expertise (Haber and Haber, 2008). Mnookin (2008) advocates a system based solely on expertise to avoid validation of ACE-V whilst also claiming that the historical successful use of fingerprint evidence, as demonstrated in cases where it has been corroborated by other evidence, has itself provided 'naturalistic' evidence that the methodology works. This approach is criticised by Haber and Haber (2008) who question the independent nature of corroborative evidence, and point out that ACE-V methodology may not have been used in all such cases. Champod (2008) argues that experts should shift from stating that they are 100% certain of their conclusions to testifying on the probability of an accurate identification and that more research should be done to provide statistical validation for the ACE-V method.

Such debate demonstrates the 'paradigm shift to a science based science' (Saks & Koehler, 2010) in certain areas of methodology validation. However, this leaves out a fundamental component of examination methodology; the examiner himself. The FBI Laboratory Committee acknowledged the subjective nature of the work of a fingerprint examiner, referring to the examiner as a 'black box' and stating that 'one may not know, understand, or appreciate the machinations that the examiner made to arrive at a conclusion' (Budowle, et al, 2006). It would seem essential that the same scientific rigour that is being employed to validate founding principles and methodology should also be applied to a better understanding and calibration of the 'black boxes' responsible for the subjective 'machinations' of fingerprint examiners. It is no longer sufficient to fall back on previous reliable results to have confidence in the reliability of this subjective decision making process, instead forensic science needs to attempt to better understand the 'black boxes' of criminalists, not just wait for the inevitable Popperian 'black swan' (Keuth, 2005).

The forensic science community is, indeed, becoming increasingly aware of the psychological issues associated with human examiners making subjective decisions. Normative psychological theory suggests a probabilistic, Bayesian type approach to decision-making, with the key assumption that the decision-maker is a rational actor who analyses decisions based upon probabilities and utilities (Hardman, 2009). It seems unlikely that this rational, economical model be said to be true of forensic scientist decision-makers, rather, that pressure and the highly emotional context of the decision may

cause the scientist to act irrationally. Kahneman et al. (1982) argue that a departure from rationality occurs due to the effect of heuristics, resulting in quicker decisions based on 'rules of thumb', or relying on external biasing information when the data present does not lead to a clear decision (Kahneman et al., 1982, Schiffer & Champod, 2007).

## 2.5. 'Cognitive forensics' – a developing field

There is a current flux of studies highlighting the presence of biasing effects within the interpretation of forensic evidence. These effects include confirmation bias (the tendency to conform to a pre-conceived hypothesis), contextual bias (being influenced by the context in which a decision is made) and belief perseverance (the tendency to accept information that supports a pre-held belief and discount information that contradicts this belief) (Kassin et al., 2013). Biasing effects have been demonstrated to occur in the disciplines of DNA analysis (Dror & Hampikian, 2011), handwriting analysis (Found & Ganas., 2013), odontology (Page et al., 2012), bite mark analysis (Osborne et al, 2014) and anthropology (Nakhaeizadeh et al., 2014). Equally, studies within the domain of fingerprinting have shown biasing effects to occur (Fraser-Mackenzie et al., 2013, Earwaker et al, 2015), and recommendations have been made to reduce such unwanted effects (Dror, 2009, Wells et al., 2013, Garrett, 2013, Dror, 2016).

The best approach to take in order to understand and compensate for failings inherent in the 'black box' examiner is currently a hot topic for debate. Views are divided as to whether the most sensible approach is to remove the need for an examiner's subjective view through relying on an objective computer generated probabilistic match determination within the pattern matching disciplines (increasing objectivity), or whether it is preferable to concentrate on exploring and correcting for the cognitive effects occurring when examiners are making these subjective decisions (improving subjective performance).

Champod (2014) highlights the recent focus upon research that investigates the effects of cognitive bias within subjective forensic science. Indeed, 'bias' can be considered to have become a popular buzz-word within research, conferences and meetings, and a motivator for organisational change within the domain. Champod (ibid) suggests, however, that a disproportionate focus of research into subjective decision making is preventing progress in increasing the objectivity with which forensic scientists can interpret evidence through a better understanding of the forensic traces themselves. Risinger et al., (2014) however, argue that there is not, in fact, an overrepresentation of research into cognitive bias within forensic science, and that research which establishes the existence and prevention of cognitive bias and that which leads to the more objective use of trace evidence need not be mutually exclusive. Buckleton et al. (2014) use the illustration of the 'human machine',



suggesting that there is a middle ground; forensic science needs to be undertaken more objectively as a result of a greater understanding of the fundamental principles of the discipline, whilst also ensuring that 'human machinery' is calibrated and its error rate is acknowledged. Buckleton et al. (ibid) propose that the existence of bias within forensic science has been sufficiently proven and that it is acceptance and action that now require further attention.

Forensic science, by nature, is a multidisciplinary domain which requires the application, with scientific rigor, of scientific disciplines to a forensic context. In recent years the forensic sciences have been criticised for not achieving this requirement and of lacking scientific integrity (Saks, 2010). With this in mind it would seem that there is indeed an imbalance in the current pool of research in relation to 'cognitive forensics'. Research that looks to increase the objectivity of the discipline through gaining a better knowledge of the fundamental properties of forensic traces and developing a statistical approach is rich in contributors from a plethora of disciplines including (but not limited to) biologists, chemists, and statisticians. Publications putting forward statistical methods often include a detailed account of the mathematical basis for these propositions (Abraham et al., 2013a). Research into the subjective aspect of practitioner decision-making, however, appears to often lack such a rigorous scientific underpinning. Whilst eminent figures from within the domain of cognitive neuroscience are figureheads for psychological research within forensic science it would appear that the majority of the underlying neuroscience behind the cognitive effects exposed through research is not often communicated to the forensic science community. It would seem that research into the subjective decisions made by practitioners would benefit from a greater theoretical background. Triplett (2013) points out that, whilst a number of solutions to the 'bias problem' have been proposed (Kassin et al., 2013), subjectivity and bias may actually be symptoms of other causes; it could be the over use of the 'bias' buzz word is limiting research which could lead to a more comprehensive understanding of the psychological mechanisms behind faulty decision making within forensic science, and so provide more effective solutions.

A greater fundamental understanding of the psychological effects that are at play is crucial to establishing the extent to which research that establishes that cognitive bias is occurring should continue. It has been suggested that there has been sufficient research to suggest that bias affects forensic practitioners (Champod, 2014) and that, as we now know that bias occurs within forensic science, further effort to identify the presence of bias within additional specialisms and processes within the domain is unnecessary. This argument implies that all practitioners carrying out all processes within forensic science are being affected in the same way by the same psychological effects. It equally implies, albeit indirectly, that the documented solutions for bias affects will be universally effective. An initial criticism to this assumption is that there is currently a lack of empirical

research demonstrating the effectiveness of any of the solutions proposed, suggesting that further research in this area is required before such solutions are widely implemented and the problem is considered to be solved. Additionally, it would seem that the only way to establish whether or not further research needs to be undertaken to establish the presence of biasing effects in novel forensic situations is to gain, through sound empirical research, a greater knowledge of the psychological effects themselves, so as to be more aware of the exact vulnerabilities of forensic practitioners. Given the differences between processes carried out within forensic science, such as the case and workplace pressures, information present, and exact cognitive processes involved, and also differences related to the same procedure being carried out in a different laboratory, or by a different practitioner at a different point in time, it would seem imperative to fully understand the effects at play in order to be able to adopt the most appropriate solution in each case, be this at a systemic or an individual level. This would also enable the continued gain of further information in relation to situations in which practitioners are vulnerable to cognitive effects in order to build up a more complete picture to inform effective targeted solutions.

As such, the present thesis seeks to carry out research which looks to investigate decision making at an early stage of the fingerprint evidential process during the fingerprint practitioner fingerprint submission decisions, and to consider this decision in terms of applicable psychological approaches.

## 2.6. Why investigate decision making within the practitioner fingerprint submission process?

'Analysis' is the first stage carried out by a fingerprint examiner when comparing two fingerprints using ACE-V methodology (SWGFAST, 2011). Analysis establishes the suitability of a fingerprint for further comparison by identifying the fingerprint characteristics present (SWGFAST, *ibid*). A study of the analysis stage found participants who had been trained in the ACE-V process able to identify more fingerprint characteristics with less intra-variability in the characteristics found than those who had not been given training (Schiffer & Champod, 2007), highlighting the difference in analysis ability between experts and novices.

Psychological studies have consistently supported the idea that a higher cognitive ability leads to higher performance in tasks (Beier & Oswald, 2012) although, there have been a number of studies that suggest that higher cognitive ability may actually be a hindrance when carrying out certain tasks (Gimming et al., 2006; Bielock & DeCaro, 2007). Ways in which experts excel and are hindered have been summarised by Chi (2006) who states that experts excel at detection and recognition of patterns but that they can be overconfident and vulnerable to bias. Dror (2012) has applied this idea to the domain of fingerprint examination, stating that expert status leads to an increased vulnerability to

bias, and focussing his research in the area to application to experts, only using novices as a point of comparison. This is problematic as there are decisions being made within the field of fingerprint evidence by 'novices'. Latent fingermarks are chemically developed by fingerprint practitioners in a laboratory, and these practitioners must then decide whether or not the fingerprint is of sufficient quality to be forwarded to a fingerprint examiner who will carry out the ACE-V process (Expert Working Group on Human Factors in Latent Print Analysis, 2012). In making this decision the practitioner is, essentially, carrying out the analysis stage of the ACE-V process but without the training that qualifies them to be an examiner, or provides expertise in this analysis process. This novice decision (relative to the analysis of the examiner) is, arguably, the most important decision in the chain of evidence as any fingerprints which are excluded at this stage will be lost. A poor decision will either lead to a waste of resources through an examiner being required to consider a fingerprint of too poor a quality, or lead to a loss of evidence.

#### 2.6.1. Potential differences in fingerprint sufficiency decision making between practitioners and examiners

Fingerprint sufficiency has been the subject of research, but this has focussed on the value assessment made by fingerprint examiners to determine whether or not to continue with a comparison. A study by Ulery et al. (2013) attempted to understand the basis for this sufficiency decision through modelling the relationships between the value determinations and annotations made by fingerprint examiners in the US. Ulery et al. (ibid) note that this process is a subjective one, with no formal criteria, which relies on knowledge and experience as opposed to a quantitative threshold. However, they found fingerprint minutiae count to be the most significant factor when determining the value of a fingerprint. Other factors were identified as the clarity of the print, the types of features present, the quantity of useable print and the relationships of the features present (Ulery et al., ibid). There is debate about the use of numerical standards by fingerprint experts throughout the ACE-V process (Evetts & Williams, 1996), with the UK no longer utilising a 16-point standard for a fingerprint match (Mackenzie, 2011). The finding by Ulery et al. (ibid) that a numerical threshold is, in practice, being used at the analysis stage may mean that there are potential inconsistencies between the decision-making capabilities of 'experts' (examiners) and 'novices' (practitioners). Langenburg (2004) found that the number of fingerprint minutiae identified by examiners was higher than that identified by lay persons, supporting the idea that experts excel at the detection and recognition of patterns (Chi, 2006, Schiffer and Champod, 2007). This may suggest that, given the same fingerprint, a fingerprint development practitioner, who has not been trained as an examiner will see less minutiae than the fingerprint examiner who has undergone comprehensive training in the ACE-V process. This demonstrates the importance of the threshold at which a development officer decides to forward a

fingerprint to an examiner, which should allow for these potential differences in pattern recognition ability.

The potential problem of differing training and expertise between examiners and practitioners is vulnerable to exaggeration by a relative absence of communication between the two domains and a lack of a joined up approach to working. This is a problem that exists at both a national and a local level in the UK. As previously described there is no national formal forum for communication between the laboratories and bureaux, and laboratories are not represented at either the Fingerprint Quality Standards Working Group or the Fingerprint Strategic Network (Stow, 2014). This means that key stakeholders within fingerprint development are not necessarily kept up to date with changes in bureaux policy and procedure, and are not represented in discussion or consultation. It would seem to be the case that the policies, procedures and working practices of the bureau are essential knowledge in order to ensure that the end product of the laboratory, the submitted fingerprint, is suitable for the requirements of the customer, the bureau. Equally there is often a lack of communication on a local level, within individual Scientific Support Units, which may result in a lack of knowledge of expectations of the customer, in particular if new techniques for uploading fingerprints to the AFIS system, or image transfer are employed, or if new methods of reporting are introduced. Such changes on a local level may alter the requirements for fingerprint images to the bureaux potential altering the threshold at which a mark may be deemed useable. Without communication of changes such as these it may be difficult for practitioners to submit fingerprints at an appropriate level.

A further exaggeration of the potential differences between practitioners and examiners may result from the external methods of training procured by the majority of UK police forces. Practitioners attending the College of Policing Fingerprint Laboratory Officer course will learn a standardised approach to fingerprint development with a limited focus on mark submission which uses standards set by the College of Policing for the purposes of the assessment (Lagden, 2014). Unless the requirement of the practitioner's own bureaux is well enforced and additional training is given to assert these standards, practitioners will be working to a different threshold to that of the examiners that they are working with. This initial training, which instructs practitioners to submit all pieces of fingerprint ridge detail to the bureau (Lagden, 2014), may have a particularly strong effect even if subsequent, different training is given once back in force, due to the effects of belief perseverance (Guenther & Alicke, 2008).

#### 2.6.2. The importance of feedback

There is a considerable body of psychological research into expertise and what it means to be an expert. It is acknowledged that a fundamental requirement in the development of expertise is receiving feedback (Ericsson & Lehmann, 1996). There are different types of feedback which target different aspects of the performance of an individual. Outcome feedback provides feedback solely in relation to the outcome of a decision, whilst cognitive feedback provides the decision maker with a measure of their cognitive processes used to come to a decision (Shepherd & Zacharakis, 2002). Cognitive feedback has been found to be more effective than outcome feedback (Shepherd & Zacharakis, *ibid*). It is unlikely that laboratory practitioners receive cognitive feedback. Whilst they may, on occasion, receive outcome feedback in relation to fingerprint submissions to the laboratory (Earwaker et al, 2015), cognitive feedback would require a knowledge of the decision mechanisms involved in fingerprint submission currently not available. Shepherd and Zacharakis (*ibid*) argue that decision aids are a useful tool for allowing cognitive feedback, particularly in cases in which outcome feedback is not readily available. This may be an approach that could be used to improve feedback within fingerprint submission. However, the effect of feedback on performance has not consistently been positive. Kluger et al. (1996) conducted a review of the effect of feedback interventions on performance, finding that early studies concentrated on 'knowledge of results (KR)' (outcome feedback). They found that there were confounding findings in relation to the effect of KR on performance, but that issues concerning research methodology and data analysis were not discussed, and positive results were selectively cited over negative findings, perpetuating the belief that KR is successful in improving performance (Kluger et al., *ibid*). The presence or absence of feedback to laboratory practitioners may be an influencing factor in practitioner decision making performance. As such it is important to be aware of current processes in relation to feedback when investigating the effectiveness of decisions made within the laboratory. It would seem that there may be differences in the decision making performance of practitioners who have closer working relationships with their fingerprint bureaux, or work from the same site, allowing more frequent feedback.

### 2.6.3. Existing research into practitioner submission decision making

Research into the effectiveness of fingerprint sufficiency decision making has focussed on examiner decision making during the ACE-V process. Langenburg et al. (2009) acknowledged that fingerprint evidence had been challenged in the case of *Minnesota vs Columbus* under the Frye admissibility criteria (that if the method is generally accepted within the community then it will be accepted in the court) on the grounds that ACE-V methodology cannot be demonstrated to be reliable. As such they identified the need for a validation, or 'method performance' study for the use of the ACE-V process by fingerprint examiners. They carried out a three phased experimental process investigating the

repeatability and reliability of the ACE-V process, finding that examiners were reasonably consistent in their judgements and finding a limited number of false positive and false negative decisions (Langenburg, 2009).

It is important to establish whether the processes used by fingerprint laboratories are effective in filtering out poor quality fingermarks that would be useless to an examiner, and retaining fingermarks of sufficient quality for an examiner to analyse. Very little research has been published exploring the practical reality of this situation.

A study by Neumann et al. (2011), however, examined the evidential value of fingermarks that were discarded within a US fingerprint laboratory, aiming to establish the cost effectiveness of this process. Examination of a sample of partial fingermarks that were discarded by laboratory staff and scene of crime officers led to a small (2.3%) increase in evidence at an estimated cost of \$138,000, suggesting that very few exploitable fingermarks are not recovered and that disproportionate investment would be required to progress these marks. However, this estimate, presumably, takes into account further processing of all discarded marks to yield a small percentage of results, not just those that would lead to an identification, suggesting that more effective decision-making could limit this cost by focussing resources at a more accurate submission threshold. Results also showed that the filtering process of submission to examiners was successful, with only a small proportion of insufficient marks being submitted by the laboratory. This would seem to suggest that fingerprint laboratories are reasonably successful in their role of submitting fingermarks to examiners that are of sufficient quality for comparison, and discarding marks of insufficient value, resulting in a system that is fairly cost effective and suggesting that there are not, in practice, significant differences in the abilities of 'novices' and 'experts' to assess the value of fingerprint ridge detail.

However, Neumann et al. (ibid) did discover that fingermarks were discarded by laboratory staff which contained very large numbers of minutiae, leading to discussion about the need to be more scientifically robust in this area of decision making. The study does not distinguish between fingermarks that were recovered by scene of crime officers and those recovered by the laboratory practitioner. There is benefit in making a distinction between these two sets of results as training, working environments, policies and procedures vary considerably between these two occupations. The results reported are based on the decisions of only one laboratory practitioner and four fingerprint examiners in a small US laboratory. Given the amount of intra-variability in novice analysis ability highlighted by Schiffer and Champod (2007), it would seem important to carry out similar research with a larger sample of laboratory staff in order to be able to generalise the results beyond this laboratory.

Earwaker et al. (2015) carried out a study that looked to assess the efficiency of the fingermark submission process within a UK laboratory, in relation to the requirements for the customer, their internal fingerprint bureau. A task was devised that allowed for the 'ground truth' in terms of sufficiency for comparison to be determined for a series of ambiguous fingermarks (those considered borderline for submission to a bureau). This involved internal examiners attempting to identify the fingermarks back to source, thus establishing whether or not they were of sufficient quality to be compared. Once this usability determination was established fingermarks were given to 11 practitioners from the Fingerprint Enhancement Laboratory of the Metropolitan Police Service who were asked to state whether or not they would submit each mark to the bureau. Responses given by the practitioners were compared to the usability determination of the examiners. The study found that 34% of the decisions made by practitioners in relation to the experimental marks were erroneous in relation to the usability determination of the examiners. Of this erroneous percentage of decisions half were decisions to submit a fingermark that was of too poor a quality for comparison (a false positive error) and half were decisions to discard a fingermark that was of sufficient quality for comparison (a false negative error) (Earwaker et al., 2015).

The present thesis seeks to build upon the preliminary work included within Earwaker et al. (2015). Firstly, this thesis seeks to establish the repeatability of the findings within another UK fingermark visualisation laboratory. This is a valuable exercise in order to establish whether such a make-up of decision outcomes is unique to the Metropolitan Police Service or can equally be applied to other UK laboratories, given that the Metropolitan Police is distinctive in its training provisions which are provided in-house instead of at the College of Police National Training Centre (Hall, 2014), and given the potential differences in structure and communication between different scientific support departments across the UK (as discussed in section 2.3). Secondly, this thesis seeks to further explore the mechanisms behind the fingermark submission decision alongside the consideration of a number of pertinent areas of psychological study.

## 2.7. The consideration of errors within forensic science

The investigation of the efficiency of the fingermark submission process requires firstly, a definition of success, and secondly, requires reference to occasions upon which success in decision making according to such a definition has not been achieved, or an 'error' in relation to the ideal benchmark has been made. The importance of discussing errors in relation to forensic science processes was highlighted within the US firstly as a result of the introduction of the Daubert Standard for admissibility of evidence (Daubert vs Merrel Dow Pharmaceuticals, 1993) and the later publication of the National Academy of Sciences Report (National Research Council, 2009). Equally, in the UK, the need for quality

standards and adherence to Codes of Practice and Conduct in order to prevent error has been recently communicated by the UK Forensic Science Regulator (Tully, 2015). However, the communication surrounding errors within forensic science can lead to misunderstandings and misinterpretations, due, in part, to confusion concerning different types of errors occurring within the forensic science process (Christensen et al., 2014). Christensen et al. (ibid) describe the four different types of error present within forensic science processes as instrument error, statistical error, method error and practitioner error. Instrument error refers to the discrepancy between the reading of an instrument and an actual or true value (requiring calibration), statistical error refers to the deviation between actual and predicted values, and method error results from traits overlapping within a group or population (error which is inherent in the task and cannot be minimised) (Christensen et al. Ibid). Whilst these errors can be seen to be of importance in the wider process of the forensic application of fingerprint evidence, it is the idea of practitioner error, defined by Christensen et al. (ibid) as “can be random or systematic, can be due to negligence or incompetence, but in most part is unintentional and unquantifiable”, which could be considered to be the most important in relation to fingerprint submission decision making. In particular, it is the ‘unintentional’ practitioner error (the subconscious error which is inherent in interpretative processes due to the way in which the human brain processes complex information) which is of most interest. Christensen et al. (ibid) state that such practitioner error can be reduced by quality assurance systems, training, and proficiency testing.

Consideration of the psychological literature in relation to errors reveals a difference in approach and mentality around the discussion of errors in an organisational setting. Frese and Keith (2015) discuss the tendency for a negative mindset and language around error making, which is indeed, commonly referred to within the forensic science literature (Cole (2005), Thompson (2010)). Frese et al. (1991) first introduced the key concept of error management to the dialogue around error. They clearly distinguish between two organisational approaches to dealing with error: error prevention and error management. Frese and Keith (2015) argue that all errors cannot be prevented due to their ubiquitous nature, partly due to the tendency of human cognition to be prone to heuristic processes (Reason, 1990). Individuals and organisations, however, have a tendency to view error making in a negative way and as an indicator of poor performance or negligence (Mangels et al., 2006), and so try to prevent these errors from occurring (Zakay et al., 2004). Frese and Keith (2015) describe error management as “effectively dealing with errors after they have occurred with the goal of minimising negative, and maximising positive, error consequences”, as opposed to the view that all errors can be prevented. Such an approach acknowledges the inherent nature of errors and looks to learn and innovate as a result of them through an open and honest dialogue. Processes in error management primarily involve the detection of errors and the reduction or avoidance of the negative consequences of this error. In



terms of fingerprint submission, error management processes could involve activities such as the assessment of a marked up exhibit by a fingerprint examiner to determine the presence of any usable fingerprints not selected for submission. The outcomes of error management processes can lead to learning from errors, improved performance, and innovation (Frese and Keith. 2015). In the example of the examiner assessment of a marked up exhibit, this could involve feedback in the mark selection process, reducing the chance of the loss of fingerprint evidence from future stages of the evidence recovery process, and lead to further dialogue around future innovations to manage potential missed marks. A move to a culture of error management instead of error prevention could also be beneficial as such an accepting and transparent culture could help to prevent hiding and blame in relation to errors.

Whilst error within forensic science and the wider legal system should, indeed, be prevented where possible, the benefit of a culture in which error can be openly discussed and managed and has the potential to lead to individual and organisation learning cannot be ignored. This is particularly the case in relation to unintentional and subconscious practitioner error occurring during interpretative perception and decision making tasks. As such, this thesis uses terminology of error to describe deviation from the established definition of decision success, but this should be considered in terms of the acceptance and subsequent transparent management of and learning from error, as opposed as an event with associated individual blame.

## 2.8. The challenge of establishing ground truth

Within forensic science the term ground truth is used to describe a trace of known source, or origin. This could entail knowing the donor of a latent fingerprint, or knowing the conditions under which this mark was deposited and stored prior to development. Knowing this information provides a factual benchmark against which experimental and procedural outcomes can be compared. For example, knowing the donor of a latent fingerprint (through recording this at the point of deposition) enables a fingerprint identification in relation to the same fingerprint to be determined to be correct or incorrect in relation to this known ground truth. As such, collections of fingerprints of known ground truth (in terms of donor and deposition) can be generated and used for research or competency testing purposes (Mikaelyan and Bigun, 2012). However, when considering the process of fingerprint submission from the laboratory to the bureau the ground truth of a fingerprint is more challenging to establish as this cannot be ascertained at the point of deposition, rather, this fingerprint must be subsequently assessed by a fingerprint examiner so as to determine its usability. Therefore, ground truth, in terms of fingerprint submission decision making, relies upon the accuracy of the usability determination of the fingerprint examiner (a subjective decision). As such, the consideration of error

and performance within fingerprint submission is more complex than in other domains in which a completely objective ground truth can readily be established. Therefore, in order to establish the effectiveness of fingerprint submission from the laboratory to the bureau it is not sufficient to consider the origin of the fingerprint itself, rather consideration must be given to the ways in which this fingerprint quality assessed by human decision makers throughout the evidential process.

## 2.9. The application of psychological theory to the fingerprint submission decision

### 2.9.1. The potential biasing effect of crime context

Within fingerprinting, current cognitive research focussing on the analysis stage of the ACE-V process has highlighted the potential for bias to occur due to the circular nature of the process often used by fingerprint experts which allows the analysis of a fingerprint to occur alongside a comparison print. Fraser-Mackenzie et al. (2013) found that when examiners were given a non-matching comparison print they were more likely to state that a fingerprint was suitable for comparison, whereas when they were given a matching comparison print they were less likely to say that the print was suitable for comparison, compared to analysing the print in isolation (Fraser-Mackenzie et al., *ibid*). Recommendations have been made by Dror (2009) that fingerprint examiners should work linearly, carrying out the analysis stage of ACE-V in isolation before being exposed to a comparison fingerprint, thus removing these endogenous biases. Wells et al. (2013) propose the 'filler control method' building upon the idea of an evidence line up in forensic identification disciplines (Garrett, 2013) which has the potential to remove contextual biases by including more than exemplar print during analysis.

With endogenous contextual bias occurring at the analysis stage of ACE-V, it would seem that there are potential benefits to a successful sufficiency decision being made by a fingerprint development officer. This may be beneficial due to the naturally linear nature of the laboratory sufficiency decision process which occurs without any exposure to a comparison print by individuals who are not trained in fingerprint comparison, limiting the effect of endogenous contextual bias on this decision.

However, studies on fingerprint examiner decision-making have also shown exogenous contextual biasing to occur. The context of a comparison print has been found to affect the analysis of the fingerprint from the crime scene (Dror et al., 2011), a fingerprint match has been shown to be more likely in an emotional context (Dror et al., 2005), the same examiners shown the same fingerprints have been shown to give a different result when the context was altered (Dror et al., 2006), and fingerprint examiners have perceived emotional context to have an influence on their analysis (Hall & Player, 2008). Charlton et al. (2010) found that decision thresholds of fingerprint examiners were vulnerable to distortion due to the effects of emotion and the need for closure. These biases may

perhaps have an even stronger effect on decisions made in a fingerprint visualisation laboratory where emotive items of evidence from crime scenes are dealt with directly. It could be suggested that this effect may vary between cases of volume crime that may illicit a lower emotional response, than serious crimes which may illicit a high emotional response, causing a difference in fingerprint submission threshold according to category of crime.

In order to establish whether any differences in submission threshold for marks related to serious and volume categories of crime can be attributed to cognitive effects, it is important to relate these findings to current fingerprint laboratory submission policy. It has been acknowledged that there is the need for a sound quantitative approach to fingerprint quality assessment (Murch et al., 2012) and, operationally, such a quantitative approach is currently lacking within UK fingerprint laboratories. With the lack of an adopted quantitative, and, indeed, objective methodology for quality assessment each laboratory must carry out its own quality assessment according to its own internal policy. It is important that the policies and procedures adopted by individual laboratories are considered when investigating the efficiency of fingerprint sufficiency decision making and the effect of contextual information on this decision making. Given the financial pressures upon modern day forensic investigation (Charlton, 2013) there is often a need to prioritise forensic work in the case of certain exhibits or indeed cases, so as to provide a best value service within the available resources. As such it would not be uncommon for a fingerprint laboratory to have a different procedure for the treatment and mark recovery in relation to volume and serious crime exhibits. An understanding of these policies and procedures is vital to ensure that an objective decision making policy is not mistaken for the effects of cognitive bias within the decision making process. It is important to be able to distinguish conscious desirable effects of context from those which are subconscious and undesirable.

Earwaker et al. (2015) found crime context to have an effect upon the fingerprint submission threshold of laboratory practitioners within the Serious Crime Fingerprint Enhancement Laboratory of the Metropolitan Police Service. The submission threshold of laboratory practitioners was found to be lower in the case of ambiguous fingerprints presented in the context of a volume crime, than those presented in the context of a serious crime. However, this laboratory routinely only dealt with serious crime exhibits, suggesting that this effect could have been due to practitioners being asked to carry out the unfamiliar task of the submission of fingerprints within a crime context in which they did not ordinarily work. As such, additional empirical investigation to assess the effect of crime type of fingerprint submission decision making within police forces who routinely deal with all types of crime (as would be the case in the majority of UK fingerprint laboratories) would be a valuable extension to this study.

### 2.9.2. The relationship between confidence and accuracy in decision-making

Confidence has been described by Kahneman and Tversky (1982) as “the degree of belief associated with what we ‘think’ will happen” (p. 515). It is a well reported cognitive effect that decision-makers tend to be overconfident in their accuracy in cognitive tasks. This effect occurs when the subjective confidence of a decision maker is higher than their objective accuracy. Traditionally within psychology the discrepancy between confidence and reality is termed calibration, meaning that overconfidence is a measure of calibration (Skala, 2008), however, calibration can also be described as risk intelligence (the extent to which an individual is aware of the risks that they are taking) (Evens, 2012). Psychologists have explained the ‘over confidence effect’ as a result of the way that people tend to focus disproportionately on explanations and reasons that provide support for their decision being correct, and consider less the evidence that suggests they may be incorrect (Harvey, 1997). Studies addressing overconfidence have been widely conducted, and a review of such literature by Fischhoff (1982) concluded that the concept of overconfidence is a robust one that is not easily removed. Overconfidence in an inaccurate decision can lead to this decision being presented with confidence, which, in turn, can cause others to be convinced that the decision is correct. The overconfidence effect can lower decision thresholds, increasing the likelihood of a poor decision. Indeed, Plous (1993, p. 217) wrote “no problem in judgement and decision making is more prevalent and more potentially catastrophic than overconfidence”.

Whilst the phenomenon of overconfidence has widely been reported, this has tended to be confined to intelligence based, cognitive tasks such as general knowledge recall. Research into perceptual (or sensory) tasks was initially limited as historically such psychophysical experiments employed vague qualitative confidence ratings which limited the potential for analysis (Baranski & Petrusic, 1994). Using a percentage scale to record confidence, Adams (1957) found under-confidence to occur in visual perception tasks. Dawes (1980) hypothesised that, whilst overconfident in their intellect, people tend to be unaware of their perceptual abilities and so are under-confident in their abilities in this area. Winman and Juslin (1993) also argue that confidence assessments are fundamentally different for cognitive and sensory tasks. It has been found that people are often under-confident when carrying out perceptual tasks involving sensory encoding (Keren, 1988, Björkman et al. 1993) and that this affect is difficult to overcome (Björkman et al., *ibid*). Within intelligence based cognitive tasks it has been found that the overconfidence effect increases with the difficulty of the task being completed. Keren (1988) found that under-confidence in perceptual tasks lessened when the tasks undertaken became more difficult, meaning that, as with the intellectual tasks, overconfidence is greater when the task is more challenging. Baranski and Petrusic (1994) also identified under-confidence in a perceptual task in which participants were asked to judge which of two lines were the closest to a

central line at various degrees of difficulty. They too found that participants were overall under-confident in this visual perception task and that under-confidence decreased (thus, there was an increase in overconfidence) as the task became more difficult. Another interesting finding of this study was that the proportion of correct responses was higher when the participants were put under pressure (through financial reward) to be more accurate than when they were put under the same pressure to work quickly.

These findings are of potential interest if considered in the context of laboratory practitioner fingerprint submission decision making. It could be argued that determination of fingerprint submission to a fingerprint examiner is both a sensory (or perceptual) task involving the recognition of patterns, but also a cognitive task in the way that the pattern of the ridge detail present needs then to be assessed against either a personal or policy driven criteria. It could be argued that the presence of either overconfidence or under-confidence in practitioner submission decision making may provide some evidence for the predominant task type being employed during fingerprint quality assessment. It may also be the case that practitioner confidence levels vary according to the difficulty of the submission decision being made. The primary changeable factor in decision difficulty would seem to be the visual (sensory) aspect of the mark itself, or possibly how easily the features present within the marks are aligned to an internal submission threshold or external policy. The combination of the two types of task involved in fingerprint submission, and the potential for over or under confidence to effect submission threshold make practitioner confidence a potentially important area of research. Equally confidence ratings within practitioners may provide an indication of the culture in which they are working. Charlton (2006) and Leadbetter (2007) report that fingerprint examiners feel they are discouraged from displaying uncertainty or self-doubt, suggesting that overconfidence may be likely to be found in the decisions of fingerprint laboratory practitioners, regardless of the type of task being undertaken, as they work in similar environments and under similar pressures to fingerprint examiners. The finding that increased motivation to work accurately rather than to work quickly had an impact upon quality of decision making (and such reduced overconfidence) (Baranski & Petrusic, 1994) further highlights the importance of working environment and organisational cultures and pressures on the output of practitioner decision making.

### 2.9.3. The application of signal detection theory to fingerprint sufficiency decision making

Signal detection theory was developed in the 1950s and is classed as psycho-physics. It was originally intended for use with radar technologies and was concerned with the difference between detecting and not detecting an enemy ship on radar (Phillips et al., 2001).

Signal detection theory (SDT) is specifically designed to deal with situations of ambiguity (Phillips et al., *ibid*). Forensic practitioners must often work within such ambiguous situations (Swets, 1992) which have found to be particularly challenging in relation to objective decision making (Kahneman et al., 1982). In such ambiguous cases SDT allows for the division of a subjective decision into two separate aspects; discrimination ability and decision threshold (Phillips et al., *ibid*). Discrimination ability is the “perceptual ability to discern similarities and differences among stimuli” (Phillips et al., *ibid*). Essentially, it is an individual’s ability to detect a signal from a background noise. This ability can be seen to be a function of the ability of the practitioner and the quality of the evidence present (Phillips et al., *ibid*). Good quality evidence could be considered to be that in which there is a high (as opposed to low) signal to noise ratio. In the case of the sufficiency determination ability of a fingerprint practitioner, discrimination ability could be defined as the practitioner’s ability to perceive the quality traits of the fingermark. This information could be considered to be a number of different things. For example, in some laboratories across the world, quality information in a latent mark may simply equate to the number of minutiae present as this would be the determining factor utilised by examiners in mark quality analysis (Mackenzie et al., 2011). As such, the ability of a practitioner to determine the presence of a minutiae from the background noise of the mark or surface would be their discrimination ability. In contrast, in countries in which quality determination is more subjective and requires the consideration of factors other than simply minutiae, discrimination ability may lie in a practitioner’s ability to distinguish ridge flow, the presence or absence of pattern type, and third level detail within the print, as part of a more subjective view of quality. Phillips et al. (2001) suggest that there is likely to be differences in discrimination ability between different forensic laboratories and variation between examiners, but that similarities in knowledge, experience, skills, and technology may lead to similarities in discrimination ability. Given the differences in training and skills between laboratory practitioners and examiners there may be difference in the discrimination abilities of the two groups. Equally differences in training as well as policies, procedure and technology may suggest differences in discrimination ability between the laboratories of different police forces.

The second aspect of subjective decision making, as defined by signal detection theory, is decision threshold. Decision threshold is the perceived line that turns a presumed negative into a positive (Phillips et al., 2001), or, in the case of fingerprint submission, the point at which a decision to submit a mark changes from being a correct submission to being a false positive. Two factors are responsible for determining this threshold in relation to each decision made; prior probability of a positive, so how often a submitted mark has been good enough for comparison in the past, and the utility value (motivation) associated with each possible outcome, in this case the perceived utility value of

submitting a poor quality print versus that of discarding a good quality print as well as those of a correct submission and a correct decision to discard.

It is the decision threshold of an individual practitioner that is vulnerable to psychological factors affecting perceived prior probability and motivation. It would also seem that it is the decision threshold of a practitioner that could be improved the most through feedback, as this would be a mechanism for gaining a more accurate view of the prior probability of a positive. Equally it is likely to be threshold and not discrimination ability that will be affected by the type of crime dealt with as this may affect the utility values assigned to each possible outcome, from a decision theoretic perspective (Biedermann et al., 2008, Gittelsohn et al., 2013).

The distinct division of decision making ability into discrimination ability and decision threshold highlights the importance of considering both of these aspects when investigating decision making performance. This has implications for proficiency and competency testing as well as, potentially, within aptitude testing. There is also an argument for the importance of determining the relative importance of the two abilities within fingerprint sufficiency decision making.

Signal detection theory (SDT) has been applied in a number of domains since its original use for military applications (Eubanks & Killeen, 1983). These include; uses within psychology to gain a better understanding of human decision making (Swets & Birdsall, 1967, Banks, 1970, Craig, 1987), and within diagnostics in the domain of medicine, including within radiology (Berbaum et al., 1989), psychiatry (Somoza and Mossman, 1991), and dentistry (Versteeg et al., 1998). Phillips et al. (2001) recommend the application of signal detection theory to decision making within forensic examination and it has been suggested that fingerprint examiners are experts in fingerprint quality signal detection (Thompson et al., 2013). As such, the application of SDT to sufficiency decision making by laboratory practitioners could be seen to be potentially fruitful.

SDT is suitable for application to fingerprint sufficiency decision making as the decision meets the requirements of analysis described by Mickes et al. (2012). Within the practitioner decision there are two states of the world (the fingerprint is sufficient for comparison, or the fingerprint is insufficient for comparison), some degree of information about the true state of the world is available (the quality features of the mark are available for appraisal), and a decision is made (the practitioner decides whether to submit the mark for comparison or discard it).

In order to determine diagnostic accuracy through the application of signal detection theory, receiver operating characteristics (ROC) analysis can be carried out (Mickes et al., 2012). ROC analysis refers to measurements of sensitivity and specificity (true positive and true negative values, respectively). In

relation to the fingerprint practitioner fingermark sufficiency decision sensitivity can be said to relate to the number of marks that are of sufficient quality and are submitted (true positives), while specificity would refer to the number of fingermarks that are insufficient and are discarded (true negatives). In addition, '1 – sensitivity' would relate to the proportion of sufficient marks discarded (false negatives) and '1 – specificity' would relate to the proportion of insufficient fingermarks that are submitted (false positives). Sensitivity and specificity calculations as applied to fingerprint sufficiency decision making are provided in Figure 2.2.

Figure 2.2. Sensitivity and specificity calculation for ROC analysis for fingerprint submission (adapted from Mickes et al, 2012)

		DECISION		
		Submit	Discard	
FINGERMARK STATE	Sufficient quality	#True Positives ( $N_{TP}$ )	#False Negatives ( $N_{FN}$ )	$N_{Positive} = N_{TP} + N_{FN}$
	Insufficient quality	#False Positives ( $N_{FP}$ )	#True Negatives ( $N_{TN}$ )	$N_{Negative} = N_{FP} + N_{TN}$

$$\text{Sensitivity} = N_{TP} / N_{Positive}$$

$$\text{Specificity} = N_{TN} / N_{Negative}$$

$$1 - \text{Sensitivity} = N_{FN} / N_{Positive}$$

$$1 - \text{Specificity} = N_{FP} / N_{Negative}$$

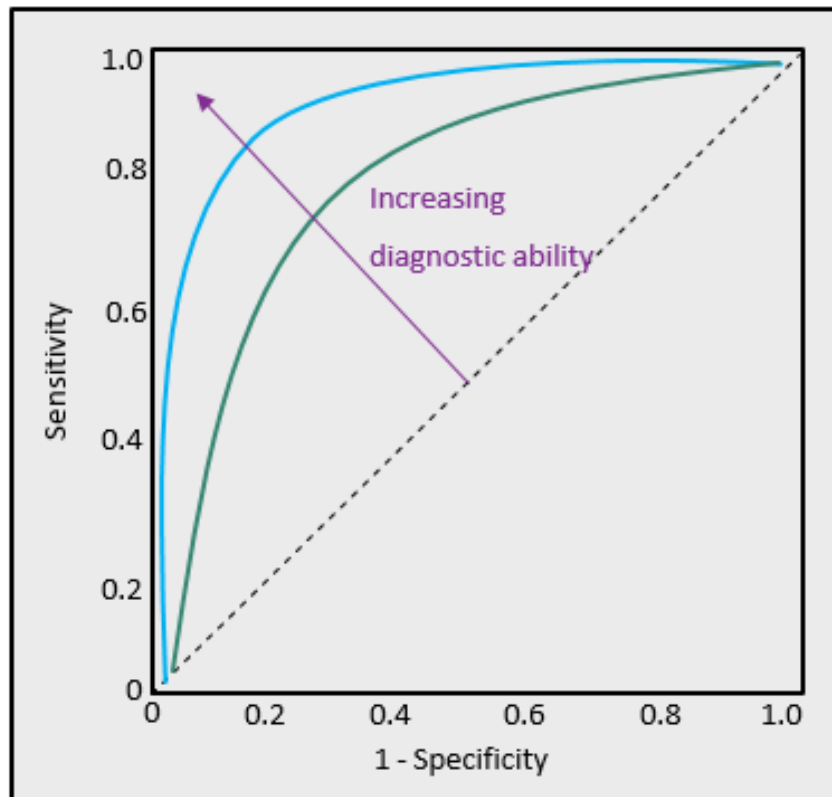
$$\text{sensitivity} = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad \text{specificity} = \frac{N_{TN}}{N_{FP} + N_{TN}}$$

A ROC curve is a series of plots of sensitivity (true positive) values against 1 - specificity (false positive) values that are associated with a single test. ROC analysis is typically carried out in order to compare the diagnostic effectiveness of different tests. Commonly, in the medical domain, these tests provide outputs on a fixed scale (for example a reading between 0 and 100). This means that cut off points from across the continuum of 0 – 100 are selected (one of these is often the signal detection threshold that is currently used), and sensitivity vs 1 – specificity is plotted for each of these points. If the results sit along a central line, then this demonstrates that the test shows no diagnostic ability whereas results at the top left of the plotted chart mean perfect diagnostic ability. If one test has greater diagnostic ability than another then the curve generated for this test will fall nearer to the ideal position (top



left). The effectiveness of each test can be quantified by measuring the area under the curve (a value of one would be gained for a perfect test), and the test with the area under the curve nearest to one can be said to have the best diagnostic ability (Mickes et al., 2012). An example ROC curve is provided in Figure 2.3.

Figure 2.3 - An example of a ROC curve



ROC curves can also be generated when the diagnostic tool is the subjective decision making of a human practitioner. Such analysis relies upon reported confidence ratings made at the time that a decision was made. Confidence ratings are treated in the same way as the diagnostic output of a test, with a number of confidence values selected and sensitivity plotted against 1 – specificity for each of these points. The position of the resultant curve can be analysed in the same way as for objective data and the area under the curve can be calculated to get a quantitative measure of diagnostic ability. This technique has previously been used within radiology in which confidence ratings are often used to facilitate decision making when establishing the presence or absence of a trace (Mickes et al, *ibid*). This approach has also been adopted in relation to assessing the diagnostic ability of eyewitnesses in identity parades (Mickes et al., 2012).

ROC analysis has the potential to be applied to fingerprint practitioner decision making in two ways. The quality sufficiency determination is subjective and non-instrument based which would suggest the possible value of using self-reported confidence ratings to determine differences in diagnostic

ability between practitioners and laboratories. The extent to which a more objective methodology can be employed depends upon the definition of quality within a fingerprint. If minutiae count could be said to be a fundamental factor in quality determination, then knowing the ground truth of minutiae count for a fingerprint could provide a benchmark for analysis. This way particular thresholds of relevance could be plotted. This could include a prescribed numerical standard for the number of characteristics required to be present in both a control fingerprint and a fingerprint lifted from a crime scene, such as the 16-point standard (now abolished within the UK (Evetts & Williams, 1996)), or any formal or informal internal laboratory submission criteria, for example, submitting when five or more minutiae are visible. Such analysis may help to enable a more objective approach to fingerprint sufficiency decision making through the setting of appropriate thresholds through the use of objective quality criteria.

#### 2.9.4. The application of Brunswik's Lens Model to the fingerprint submission process

The application of Brunswik's Lens Model (Brunswik, 1952) as described by Hammond et al (1964) can help to describe the fingerprint submission decision in terms of an attempt to mirror quality indicator cues between practitioners and examiners (Kahneman and Klein, 2009). In the traditional application, the model is used to describe the process that leads to a judgement on a problem that has a true solution or value, for example, determining the age or the profession of an individual of whom you have no prior knowledge. Brunswik's Lens Model describes this uncertain judgement being made through the evaluation of the weight of a number of cues. For example, in an attempt to judge the age of a stranger we may consider certain cues such as clothing style, hair colour, and signs of skin aging. We may also place differing values on each of these clues, for example, we may consider signs of skin aging to be a more powerful indicator of age than hair colour. Proportionate consideration of these weighted factors will lead to a judgement of the age of the stranger. Within the traditional Lens Model there are five key components; a criterion (for example, age of a stranger), cues (for example, hair colour), ecological validities (the reliability of the hair colour cue), cue utilisation (the weight placed upon hair colour information in the age decision), and a response or judgement (age determination) (Newell & Shanks, 2014). Fingerprint submission decision making fits well into Brunswik's framework with one major adaptation; the judgement being made is not to determine a true solution or value existing within the world, it is, rather, to successfully determine (and mirror) the subjective judgement of another person (in this case the usability determination of a fingerprint examiner). This scenario differs from typical models and theories of group decision making (such as game theory) which describe competitive decision making (Robin, 1993), and places the emphasis on

the practitioner decision as that which needs to utilise the appropriate cues and weights of these cues in order to make a submission decision appropriate to that of a fingerprint examiner.

Previous research within the domains of crime and forensic science has suggested possible differences between the use of decision cues by fingerprint examiners and practitioners. Baber and Butler (2012) compared the concurrent verbal protocols of novice and expert crime scene examiners in a scene processing task. They found that whilst novices concentrated upon dealing with the scene in terms of its features, the experts were concerned with the likely actions that could be performed as a result of the examination (Baber & Butler, *ibid*). Garcia-Retamero & Dhamai (2009) investigated the use of decision cues when determining which of two properties were most likely to be burgled. The expert group consisted of both police officers and burglars, whilst the novice group was made up of students. The study found that those within the expert group differed in the cues that they considered important in their decision making. The cues of the police officers within the expert group were actually more similar to those of the novice group than those of the burglars within the group. The findings of these two studies suggest that there could be differences in the decision cues used by practitioners and examiners, as examiners (who could be considered to be experts as, like the burglars studied by Garcia-Retamero and Dhamai (2009) they actually carry out the practical process of mark quality assessment that is required prior to carrying out a comparison) may consider cues more in terms of what could be done with the mark within the bureau such as exclusion, comparison, or search on an AFIS.

The identification of the key cues utilised during the mark quality assessment process by both practitioners and examiners is crucial to understanding the mark submission decision. The use of Brunswik's Lens Model as a basis for consideration of this decision allows the integration of the idea of cue matching between the practitioner and examiner, the identification of cues within the environment of the fingermark decision, and the assessment of the relative importance of these cues.

#### 2.9.5. Visual perception within fingermark quality assessment

Forensic identification evidence which relies solely on the subjective comparison of two patterns is psychologically problematic as when a physical stimulus (the fingerprint of a suspect) needs to be compared with a putative second instance of the same stimulus (a fingerprint from a crime scene) one or both instances of observation can be corrupted by perceptual, memorial or judgemental noise (Busey & Loftus, 2007). This is also problematic within the of field eyewitness testimony where the witness is exposed to suggestive information after witnessing an event, which is used to reconstruct or supplement a poor memory of the suspect during an eye witness line up (Busey and Loftus, *ibid*). The process of identifying an individual from a fingerprint is a judgement also based upon the

perceived similarity of two images (Busey & Loftus, *ibid*) that is arguably susceptible to 'cognitive contamination', as termed by Kassin et al (2013).

Whilst fingerprint quality assessment as carried out by a laboratory practitioner does not involve the comparison of two stimuli, the visual assessment of a fingermark in isolation could still be vulnerable to such 'cognitive contamination'. Psychology has attempted to provide an explanation for the process by which the sensory inputs received by the sensory organs is converted into perceptions of that sensory input. In an attempt to do this, views have been divided between two key, and fundamentally different, approaches differing in the extent to which human perception relied directly upon the information present in the stimulus itself, versus the extent to which perception draws upon the expectations, experience, and prior knowledge of the perceiver. Gibson (1966) suggests an ecological, or direct, theory of perception which considers perception as a bottom up data driven process which occurs through the one directional transfer of information from the retina to the visual cortex. According to Gibson (*ibid*) this process relies upon innate evolved mechanisms, and does not require any aspect of learning or hypothesising. Gregory (1970), on the other hand proposes a top down theory of visual processing. His theory suggests that contextual information and the experience of an individual aid in the understanding of a visual trace. Visual perception, according to Gregory (*ibid*) is a hypothesis based approach with the formation of incorrect hypotheses leading to errors in perception. The constructionist theory of Gregory (1970) indicates that individual differences may be observed where visual perception is used as a diagnostic tool, due to individual differences in learning and experience, as such, the absence of a comparison fingerprint during laboratory practitioner does not prevent difficulties in visual perception occurring.

#### 2.9.6. Effect of background context

##### 2.9.6.1. Decision by sampling

Many models of judgement and decision making are based upon the process of sampling from, and comparison with, exemplars, either from the decision context or from working memory. Examples include norm theory (Kahneman & Miller, 1986), decision field theory (Busemeyer & Townsend, 1993), support theory (Tversky & Koehler, 1994), and the stochastic difference model (Gonzalez-Vallejo, 2002). The decision sample used combines values from the immediate context in which the decision is being made and memories of previously acquired mark quality ratings from memory to result in a mean value upon which a comparison is made. Decision by Sampling (DbS) (Stewart et al., 2006) draws upon the comparative essence of these theories but, instead, assumes that individuals arrive at subjective attribute values through binary, ordinal comparisons (determining a two-way rank) with

attribute values from the immediate context in which the decision is made and from memory, and then ranking their decision within the available sample.

Application of this approach within the context of laboratory fingerprint submission would mean that the decision sample used to make a submission decision in relation to a particular mark (the 'target mark') would consist of the quality of marks surrounding the 'target' mark on a crime scene exhibit and the quality of marks encountered whilst making previous submission decisions. DbS assumes that decision making is based upon the use of three cognitive tools: binary, ordinal comparison, sampling, and frequency accumulation, all of which decision makers have been shown to excel at, within the judgement and decision making literature (Stewart, 2009). Underlying this theory is the principle that psycho-economic functions (the introspection that allows the decision maker to understand the decision process) are removed from the psychological process of choice: "psycho-economic functions are revealed from the choice data: they describe the choices people make, but not the psychology of choosing" (Stewart, *ibid*).

Stewart et al (2008) provided an extension of the decision by sampling model. Within the model it is assumed that decision makers make a series of binary, ordinal comparisons between attribute values in working memory and that frequency accumulators tally the number of favourable comparisons for each option. When the difference in tallies between the two options reaches a certain threshold, a decision is made. In addition, the model makes the assumption that sampling from the immediate context and sampling from memory were equally likely. Stewart (2009) suggests that there are three key stages to the process of decision by sampling: selecting a target attribute, selecting a comparison attribute, and carrying out a binary ordinal comparison. In terms of the fingerprint submission decision the two decision options (or prospects) available to the practitioner are to either submit or not submit the fingerprint. Attributes of each of these prospects can be considered to be features of the mark which indicate quality. Once a quality attribute has been selected (for example the quality of ridge flow in the mark) then a comparison attribute is selected. This may be the same quality indicator in relation to either a mark in the nearby context or in a mark in working memory. A comparison between these two attributes is carried out and the winner (the highest quality ridge flow, in this case) is determined. Comparisons according to a number of quality attributes would be made until the difference in scores reaches a certain threshold. At this point a decision would be made to submit the mark (as the quality attributes associated with it have 'won' against those of the exemplar) or to not submit the mark (as the quality attributes have 'lost' against those of the exemplar). Decision by sampling, therefore, suggests the potential importance of two key aspects in fingerprint submission decision making: the determination of quality attributes, and the selection of a comparison target (be it from working memory, or present within the context of the target mark).

#### 2.9.6.2. Contrast and assimilation effects

Within social psychology literature the effects of contextual information within evaluative judgements (such as determining the quality of a fingerprint) have been described as either contrast effects or assimilation effects, with the interaction between these effects framed within the integrative inclusion/exclusion model (Bless & Schwarz, 2010).

The inclusion/exclusion model describes an evaluative judgement as involving both a mental representation of the target stimulus (in this case an ambiguous fingerprint) and a mental representation of a standard (the context stimulus) to which this target stimulus is compared. Both of these aspects of the evaluative judgement are context sensitive and are formed based upon the information which is most accessible at the time, similarly to decision by sampling. The effect that the context stimulus (or perceived standard) has upon a judgement made in relation to the target stimulus does not, however, depend upon the ranking of the target stimulus in relation to the context (as it would within decision by sampling theory), instead the effect of the context stimulus depends upon the way in which this information is utilised by the judge. If the context information is used to inform the judge's representation of the target, then an assimilation effect will take place. For example, if a good quality fingerprint is considered next to a target mark and the information of the perceived good quality of the context mark is used to inform the judges' decision of the quality of the target mark, this will mean that the quality judgement of the target mark aligns with (assimilates) the quality of the context mark, in this case leading to a 'good quality' judgement in relation to the target mark. Alternatively, if the context information is used to inform the representation of the comparison standard then a contrast effect will take place. For example, if the information in relation to a context mark is used to inform a standard by which to compare the target mark then a good quality context mark will lead to a poorer quality determination in relation to an ambiguous target mark.

Which of these two processes occurs is considered to be dependent upon a judges' belief on the relevance of the context information, information which is perceived as representative of the target, and norms which influence the perceived usefulness of the information (Bless & Schwarz, 2010).

#### 2.10. Summary of the objectives of this thesis

A review of relevant literature has introduced the roles and relationships of the key stakeholders within the UK fingerprinting process and has demonstrated the procedural importance of the transfer of visualised fingerprint evidence from the fingerprint enhancement laboratory to the fingerprint bureau. Consideration of the operational process of mark submission and the empirical literature concerning aspects of the ACE-V process and fingerprint quality assessment tasks carried out by fingerprint examiners has led to a number of key observations that may impact upon the fingerprint submission decision making of fingerprint laboratory practitioners.

Consideration of the current fingermark submission process has identified a number of differences in the training and routine job roles of fingerprint laboratory practitioners and fingerprint examiners. Examiners were found to receive training that focusses upon carrying out fingerprint quality assessments, whereas laboratory practitioner training was found to focus, instead, upon the chemical enhancement of marks, with the national training facility used by the majority UK police forces (the College of Policing) found to recommend that practitioners submit all fingermark ridge detail to the bureau, therefore not including aspects of fingermark quality assessment within their professional qualification. Equally, within routine casework, there was found to be the opportunity to receive feedback on the outcome of fingerprint examiner mark quality assessment decisions, whereas it was identified that there was a lack of a mechanism by which laboratory practitioners could receive feedback in relation to their decision not to submit a mark, in terms of the opinion of a fingerprint examiner. In addition, a historical lack of communication between fingerprint laboratories and bureaux in the form of joint working groups at a national level suggests the potential for a lack of a joined up approach to fingerprint submission procedure between the two entities.

Consideration of the findings of empirical study investigating aspects of fingerprint quality assessment has illustrated such research to focus primarily upon the use of the ACE-V process (in particular the 'analysis' stage) by fingerprint examiners, tending to neglect the application of such empirical research to fingerprint laboratory practitioners. Findings of empirical study in relation to the 'analysis' process of fingerprint examiners have identified a high level of inter-examiner variation and also differences in the 'analysis' ability of examiners and novices. These findings suggest the value of further empirical study of the submission decision of practitioners, which can be considered to be similar to the 'analysis' decision of examiners, in that it is a subjective quality determination of a fingermark. This is because there may be similarities between laboratory practitioners and the novice groups used in such studies, as the expertise literature suggests that practitioners lack the feedback mechanisms that may be needed to develop expertise in this area.

Whilst one empirical study (Neuman et al., 2011) was found to consider the process of mark progression and found submission from the laboratory to the bureaux to be an efficient process, this was carried out at a very small scale within a US laboratory. Initial research by the present author sought to establish the efficiency of the fingermark submission decision within the UK MPS laboratory (Earwaker et al, 2014). The finding of inefficiencies during this process highlight the need to carry out further empirical research to both establish the reproducibility of these findings within other UK laboratories (as training is unique within the MPS laboratory), and also to further investigate how these decisions are being made.

As the fingermark submission decision relies upon subjective human decision making, literature in relation to decision making as applied within forensic science and within fingerprinting has been introduced within this literature review. This literature was found to focus upon aspects of 'cognitive bias' and the current debate around the focus of such research and the need for further research in this area was outlined. A paucity of such research into the factors affecting the submission decision making of practitioners was identified.

Consideration of the present literature suggests the importance of further examining fingermark submission decision making in terms of the efficiency and success of the process but also in terms of mark submission as a human decision making process. Existing literature in relation to cognitive biases in relation to fingerprint examiners suggests that practitioners may also be vulnerable to such effects due to the pressured environment and contextual information that they are exposed to. This thesis, however, seeks to examine the decision making process of fingermark submission at a more comprehensive level than establishing the presence or absence of cognitive biases.

Literature in relation to a sample of relevant psychological approaches to the study of human judgement and decision making have been introduced. These approaches and theories, and their application in other domains, highlight the importance and potential benefit of considering the fingermark submission decision from a more psychological view point. The psychological literature reviewed highlights a number of effects that have the potential to occur within, or have an influence upon mark submission. The potential biasing effect of crime context, the idea that experts in a task can be overconfident in their decision making, the consideration of a decision in terms of cues and thresholds, considering the submission decision as a cue matching task between practitioners and examiners, potential individual differences due to top down visual perception, and the effect of the background environment (such as the quality of background fingermarks) on a submission decision are all introduced.

Based upon this review of the current literature this thesis seeks to meet the following summarised objectives through novel empirical research set out within four experimental chapters, as set out in Table 2.1.



Table 2.1 - Summary of thesis key objectives and current knowledge gaps

Chapter	Key Objectives	Knowledge Gap Addressed
Chapter 3: An investigation of fingerprint submission decision making within UK police forces	Objective 3.1 – Assessment of the efficiency of fingerprint submission decision making within Laboratory A (a UK metropolitan laboratory procuring training from the College of Policing)	A lack of knowledge of the efficiency of the fingerprint laboratory submission process within a UK police force that procures training from the College of Policing
	Objective 3.2 – Assessment of the effect of contextual information on submission decision making within Laboratory A	No knowledge on the effect of contextual information on mark submission in a laboratory routinely dealing with all crime.
	Objective 3.3 – Assessment of the relationship between self-reported confidence levels and decision accuracy within Laboratory A	A lack of consideration of the role of expert performance and confidence within fingerprint laboratory practitioners
	Objective 3.4 – An assessment of inter-laboratory consistency in fingerprint submission decision making	Allows an extension of preliminary study outlined within Earwaker et al (2015), comparing the mark submission process within MPS with that of another UK police force
Chapter 4: An investigation of self-reported fingerprint quality assessment decision rationale	Objective 4.1. To explore the rationale behind laboratory practitioner fingerprint submission decision making within Laboratory A.	No published research on the fingerprint submission decision rationale of laboratory practitioners
	Objective 4.2. An investigation of inter-laboratory consistency in practitioner decision rationale	A lack of existing research looking at the consistency of reasons to submit or discard fingerprints between laboratories
	Objective 4.3. Exploring the relationship between practitioner submission decision rationale and fingerprint examiner usability rationale	No existing research comparing the usability decision rationale of examiners with the submission decision rationale of practitioners
Chapter 5: A comparison of examiner and practitioner minutiae counts, and an assessment of inter-practitioner variability	Objective 5.1 – an examination of inter-practitioner variability in fingerprint minutiae count	No published research establishing variation in laboratory practitioner minutiae counts
	Objective 5.2 – a comparison of the variability of practitioners and examiners	No existing comparison of the variation of minutiae counts within practitioners and examiners.
	Objective 5.3 – a comparison of practitioner and examiner minutiae counts	No current research establishing the minutiae detection abilities of practitioners and examiners in relation to the same fingerprints
	Objective 5.4 – a comparison of variability according to fingerprint quality	No existing examination of individual or organisational differences in minutiae count
	Objective 5.5 – an assessment of inter-laboratory differences in minutiae count and inter-practitioner variability	
Chapter 6: The effect of background mark quality on mark submission decisions	Objective 6.1. - An investigation of the main effects of background mark quality on target mark submission	A lack of existing research investigating context effects within the simultaneous presentation of fingerprints
	Objective 6.2. An assessment of individual differences in mark context effects	
	Objective 6.3. Assessing for the influence of demographic factors and the presence of order effects	

# Chapter 3 An investigation of fingerprint submission decision making within UK police forces

## 3.1. Introduction

### 3.1.1. The potential for inefficiencies in the fingerprint laboratory mark submission process

Fingerprint visualisation laboratories play a key role in enabling the transfer of fingerprint evidence from a crime scene to court. The fingerprint laboratory technician acts as a gate keeper selecting which of the fingerprints chemically developed in the laboratory are of sufficient quality to be submitted to a fingerprint examiner for comparison against fingerprints from suspects of interest. (Earwaker et al. 2015, Forensic Science Special Interest Group, 2014). The intended outcome of this decision making process is to prevent the submission of fingerprints that are of too poor a quality for comparison whilst ensuring that all fingerprints that could be compared are submitted, in order to ensure the optimum allocation of resources in examiner time and the resources associated with image capture, whilst preventing the loss of evidence of value. In order to make this submission decision accurately the quality judgement made by a laboratory practitioner would need to mirror the judgement made by an examiner during 'analysis' stage of ACE-V. It could, however, be argued that the most favourable situation would be submission at a slightly lower threshold than that employed by an examiner so as to provide a safety net to ensure that all fingerprints that are of sufficient quality to be evidentially useful are forwarded at the cost of some insufficient fingerprints also being forwarded, given that the utility value of ensuring that all evidential prints are submitted could be considered to be higher than the cost of processing a small number of insufficient fingerprints.

Differences in the focus of the initial and continued training of fingerprint examiners and fingerprint practitioners (Lagden, 2014) may suggest differences in their abilities to quality assess a fingerprint between the two groups. The College of Policing national training centre, for example, provides training to fingerprint examiners which focusses on the ACE-V process of comparison including 'analysis' of the quality of a mark, whereas Fingerprint Laboratory Officer (FLO) training run by the same institution focusses on the selected and practical application of chemical techniques for visualising latent marks and image capture of these marks (Lagden, *ibid*). The lack of an objective methodology or standard for fingerprint submission, and the lack of a nationally consistent approach means that the College of Policing instructs trainee practitioners to submit all fingerprint ridge detail as part of the assessment criteria of the FLO course, and does not acknowledge the use of submission

thresholds. The reality, however, of submitting all pieces of the ridge detail is quixotic as Scientific Support Departments struggle under resource constraints (Charlton, 2013) and time pressure meaning that, once back in force, practitioners will be required to be selective about the ridge detail that they submit to the bureau, according to any local policy in existence (Forensic Science Regulator, 2016). Subsequently, while there is a unified approach to examiner training and sufficiency criteria there is not such consistency in this area from a laboratory perspective. This difference in training and approach may suggest differences between the sufficiency, or quality, judgements of practitioners and examiners within a force, and between practitioners working for different forces.

Research has shown training to be important within fingerprint analysis. Participants given training in this area were found to identify a higher number of minutiae within a print than those who had not received training (Schiffer & Champod, 2007), suggesting that there may be a difference in signal detection ability (Phillips et al, 2001) between practitioners and examiners, meaning that practitioners may not be forwarding some fingerprints of evidential value.

However previous research in an operational setting has shown the evidence filtering processes carried out by practitioners to be predominantly successful within a laboratory in the United States, and suggested that a reduction in the laboratory submission threshold would not provide added value (Neumann et al., 2011).

### 3.1.2. Expanding the scope of previous research

#### 3.1.2.1. The effect of training and culture

Previous research (Earwaker et al. 2015) investigated the efficiency of fingerprint sufficiency decision making within the Fingerprint Development Laboratory of the UK Metropolitan Police Service (MPS). This paper is provided for reference in Appendix B.1. This research identified that there were discrepancies between laboratory practitioner and examiner sufficiency decision making, leading to 33.6% of fingerprints discarded being of useable quality, and 34.1% of fingerprints submitted by the laboratory being of too poor a quality to be used (Earwaker et al. *ibid*). These discrepancies appeared not to be occurring as a result of a straight forward difference in suitability threshold, as it was not simply the case that practitioners were just submitting poor quality marks (adopting too low a threshold), or just discarding good quality marks (adopting too high a threshold), but rather that they were making a combination of these erroneous decisions (Earwaker et al., *ibid*). Results showed that accuracy of decision making varied according to the fingerprint in question. This variation in outcome according to fingerprints of a similar level of quality may suggest the importance of the detection and appraisal of specific features or cues within each fingerprint rather than the use of a higher level threshold. Such a process of the detection and appraisal of cues takes place during the 'analysis' stage

of the ACE-V process. This process could be considered to be more challenging in the case of more ambiguous marks where the detection of quality indicators is more difficult. Given the findings of Schiffer and Champod (2007), the suggestion that signal detection is important would suggest the importance of considering differences in 'Analysis' training to be potentially an influencing factor in sufficiency decision making.

Training for laboratory practitioners within the Metropolitan Police Service is provided internally by the Metropolitan Police Crime Academy, as opposed to the Nationally Training Centre of the College Policing (Hall, 2014). As such, there may be differences in the level or type of 'analysis' training provided to Metropolitan Police Service Fingerprint Laboratory Practitioners, to that provided to other practitioners working within forces procuring their training from the College of Policing. The present study, therefore, seeks to investigate the efficiency of the mark submission process within such a police force (referred to as Force A or Laboratory A throughout this chapter). Laboratory A was selected for use in this study over other UK laboratories procuring their training from the College of Policing as it was an example of a large metropolitan laboratory where it would be possible to recruit a similar sample size of participants as had been recruited from within the Metropolitan Police Service, in addition to the important differences between the two laboratories in training procurement. There are, however, additional differences between the Fingerprint Laboratories of the Metropolitan Police Service and Force A. Whilst both laboratories carry out casework for a large metropolitan area, the Metropolitan Police Service has a number of fingerprint laboratories, whereas Force A utilises one laboratory for all its fingerprint work. The Metropolitan Police Serious Crime fingerprint laboratory has gained UKAS accreditation to ISO17025, whereas Laboratory A (at the time of participation in the present study) was working towards accreditation to meet the 2015 deadline. A further, fundamental, difference between the two forces is that the Force A fingerprint laboratory and bureau are situated on the same site, whereas the MPS serious crime laboratory and fingerprint bureau are located separate sites. Given the differences in 'analysis' training it may be the case that there are differences in the ability of the practitioners to determine the presence of the characteristics of a print that are indicative of its quality, resulting in a difference in quality signal detection between the two forces. Equally the differences in the relative location of the laboratories and bureaux may suggest the potential for different collaborative working strategies between these units which may result in differences in expected mark sufficiency threshold. In addition, it may be that differences in geographical location may have an impact upon the quantity and type of feedback mechanisms in place from the bureau to the laboratory which may affect the ability of practitioners to gain expertise in their role of sufficiency decision making, given the importance of feedback in the development of expertise (Ericsson & Lehmann, 1996).

As such, it would seem beneficial to establish whether the findings of Earwaker et al. (2015) are replicated within a different metropolitan fingerprint department (that of Force A), or whether differences in structure, procedure, feedback, and training have an effect on the efficiency of fingerprint laboratory sufficiency decision-making.

#### 3.1.2.2. The role of crime context

Contextual information has been found to affect the judgements made by fingerprint examiners (Dror et al. 2005, 2006, 2011, Charlton et al. 2010). In addition, Earwaker et al. (2015) found the threshold of laboratory practitioner fingerprint sufficiency determination to be lower in the context of a serious crime and higher in the case of a volume crime. However, these results were found in a laboratory which routinely only dealt with serious crime, and, as such, did not have experience through case work of dealing with the submission of fingerprints in volume cases or gaining feedback from a fingerprint examiner in these cases. It would, therefore, be interesting to assess the effect of crime context of fingerprint sufficiency decision making within a laboratory that routinely deals with fingerprints related to both serious and volume crime, so as to establish the robustness of the contextual effects previously observed within laboratory practitioners (Earwaker et al., 2015).

In order to gain a greater understanding of fingerprint sufficiency decision making the preliminary study carried out by Earwaker et al. (2015) within the Metropolitan Police Service was repeated within Force A. The repetition of this study allowed for observation of the efficiency of, and effect of crime context on, fingerprint sufficiency decision making within Force A fingerprint laboratory, as well as a comparison with the efficiency and effect of crime context previously found with the Metropolitan Police Service (Earwaker et al., *ibid*).

### 3.2. Summary of objectives

The present study aims to establish whether fingerprint laboratory practitioners are successful in submitting fingerprints that are useable by fingerprint examiners within a police force which is representative of the majority of UK police forces in terms of its procured training (Laboratory A). The study aims to then compare these findings to those reported within the Earwaker et al. (2015) in relation to the UK Metropolitan Police Service. This will be achieved through addressing the following research objectives:

Objective 3.1 – Assessment of the efficiency of fingerprint submission decision making within Laboratory A

To establish the extent to which fingerprint laboratory fingerprint submission decisions match the usability determinations of fingerprint examiners.

Objective 3.2 – Assessment of the effect of contextual information on submission decision making within Laboratory A

To test the hypothesis that case contextual information will affect the fingerprint submission decision making thresholds of practitioners, resulting in lower thresholds in the case of serious crime contextual information and higher thresholds in the case of volume crime contextual information (as was reported in Earwaker et al (2015)).

Objective 3.3 – Assessment of the relationship between self-reported confidence levels and decision accuracy within Laboratory A

To test the hypothesis that laboratory practitioners will be overconfident in their fingerprint submission decision making, given that decision makers tend to be overconfident in their accuracy in cognitive tasks and focus upon information supporting their decision being correct (Harvey, 1997).

Objective 3.4 – An assessment of inter-laboratory consistency in fingerprint submission decision making

To compare fingerprint submission decision making between Laboratory A and the Metropolitan Police Service Serious Crime Fingerprint Laboratory (as documented in Earwaker et al, 2015). Inter-laboratory consistency will be assessed through the following sub-objectives:

Objective 3.4.1: Comparison of the overall efficiency of fingerprint sufficiency decision-making

Objective 3.4.2: Comparison of the effect of contextual information on sufficiency decision making

Objective 3.4.3: Comparison of the relationship between self-reported confidence and decision accuracy

### 3.3. Method

#### 3.3.1. Overview of method

This study required laboratory practitioner participants to provide a submission decision in relation to a series of borderline quality fingerprints of known fingerprint examiner usability. Through doing so a determination of whether each of these decisions was in agreement or erroneous in relation to the judgement of fingerprint examiners could be made. Whilst completing the mark submission experimental task practitioners were also asked to state how confident they felt in making each submission decision and to state any reasons for this decision. Alongside this empirical task the manager of the laboratory was interviewed so as to gain information in relation to any policies governing fingerprint submission that may have impacted upon data analysis. For example, had the

laboratory manager reported a difference in submission threshold according to the context of the case documented in policy, then this would have impacted upon the way in which the case context data was discussed.

### 3.3.2. Method for objectives 3.1 – 3.3

#### 3.3.2.1. Materials

##### Development of experimental fingermarks

A set of experimental fingermarks compiled for previous research investigating fingermark sufficiency decision making within the UK Metropolitan Police Service (Earwaker et al, 2015) was utilised within this study. To produce this set of fingermarks a series of latent fingermarks of known source (digits of the right hand of the researcher) were deposited on 'clean' sheets of white A4 paper (Zerox Performer A4 laser, copier and inkjet 80g/m<sup>2</sup>) at a range of pressures and with a range of movement. A combination of naturally deposited marks and marks deposited using a Latent Print Reference Pad: Amino Acid Based™ (Lightening Powder, part no. 1-2791) were used. Marks were developed at the UKAS accredited Metropolitan Police Fingerprint Laboratory Lambeth, London with the use of the reagent Ninhydrin to visualise the amino acid constituents in the deposited latent fingermarks, as per Home Office Centre for Applied Science and Technology recommended, and ISO1725 accredited, procedure for fingerprint development on smooth porous items (Bowman, 1998). Each paper was submerged in pre-prepared Ninhydrin Working Solution (Samuel Banner & co. Ltd., mixed on 10-04-13) and left to dry within a fume cabinet. The papers were then placed in a Ninhydrin Oven (Weiss Galenkamp, calibrated 7/3/13) set at 80°C and 65% RH. The door was closed and the oven was allowed to regain its regulated temperature. The papers were removed from the oven after two minutes, placed in a plastic folder to prevent contamination and left for six days to allow any further mark development. Forty areas of developed ridge detail considered borderline in quality for submission to an expert were selected by the researcher, with advice from a Metropolitan Police Fingerprint Laboratory Practitioner. These areas of ridge detail were each identified with a sticker (labelled from A-AN) and a fictitious case number, as per standard Metropolitan Police 'Marking Up' procedure. The selected areas of ridge detail were photographed in colour using Metropolitan Police standard mark photography procedures, and were printed 1:1 on photographic paper.

##### Examiner assessment of fingermarks

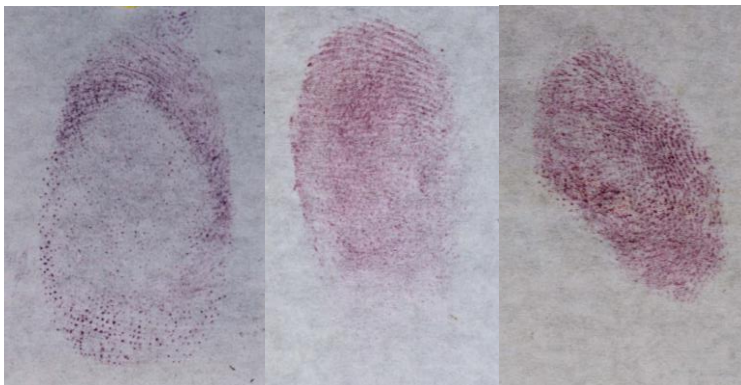
Inked fingerprints of the same source as those developed were taken by two Metropolitan Police Fingerprint Examiners. Three sets of prints (termed sets of tenprints within UK fingerprinting) were taken from each hand, including all areas of ridge detail. Prints were recorded under an assumed name.

Five fingerprint examiners assisted with the set-up of the study. Two were employed in the Metropolitan Police Serious Crime Bureaux, two were employed in the Fingerprint Bureau of Police Force A, and one was a registered independent examiner. Each examiner was given a set of the tenprints described above and one set of printed mark photographs. They were asked to attempt to identify each of the 40 mark photographs to the tenprints, to make a general observation of the quality and clarity of the mark, to state whether or not the mark was searchable and comparable, and whether they were able to identify the mark (if so to which finger), or whether it was inconclusive, excluded or insufficient. Each examiner worked independently resulting in five independent examiner opinions of the quality of each of the 40 fingermarks. See Appendix B.2 for completed examiner record sheets.

#### Selection and presentation of experimental fingermarks

20 of the 40 examiner assessed marks were selected to form the 'experimental image set'. The opinions of the five examiners were consistent in the comparability of the marks and whether or not they could be identified, with the exception of only one mark about which the independent examiner disagreed with the judgements of the four in-force examiners. The set represented an even mixture of ten insufficient and ten identifiable marks as stated by the examiners, with a range of perceived quality. See tabulated profile of marks selected in Appendix B.3. An example of three experimental marks used in the study is provided in Figure 3.1.

Figure 3.1. Example experimental fingermarks (not to scale)

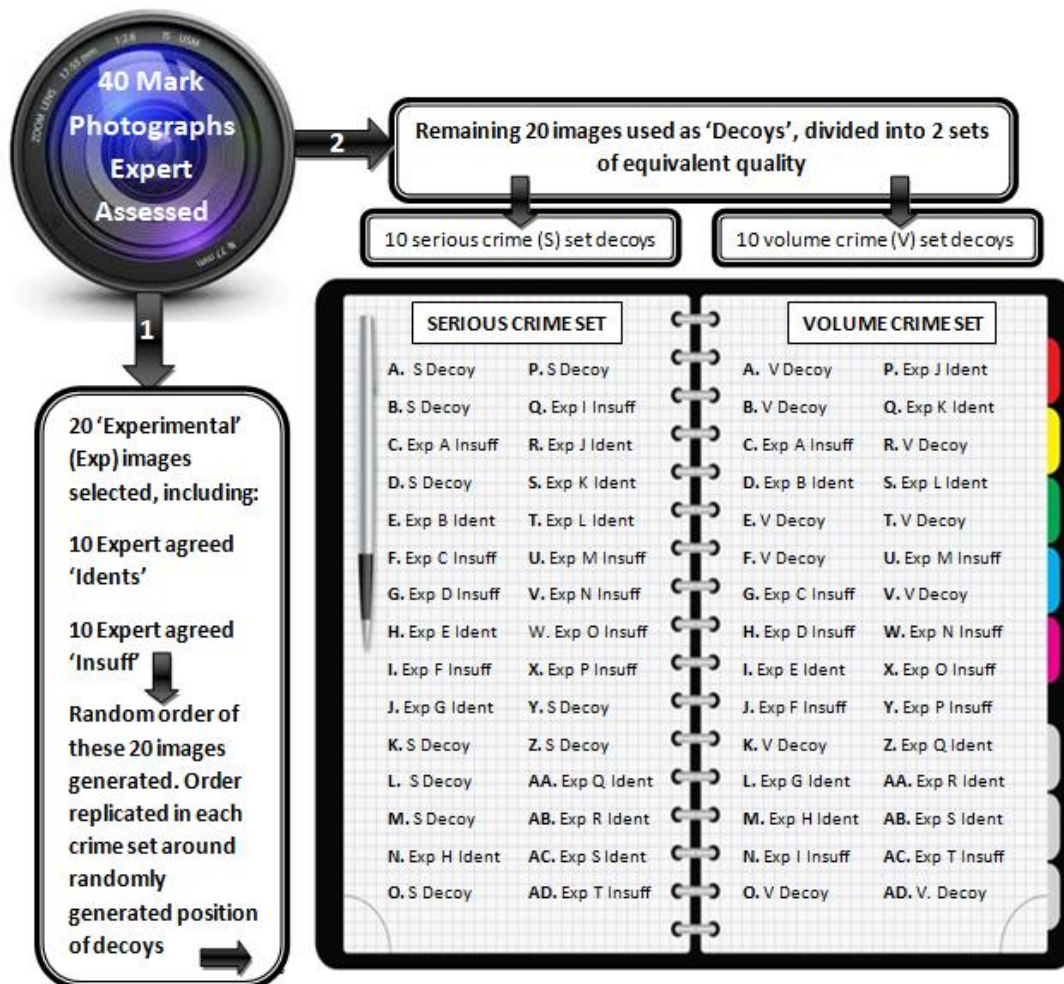


The remaining fingermark images not selected to form the experimental set were divided into two sets of ten 'decoy prints' ('set 1' and 'set 2') to be included among the experimental images. Agreement by all five experts was not necessary for these images as the results of these were not to be included within the data set for analysis. Tabulated details of the decoy marks can be seen in Appendix B.4. The experimental image set and decoy set 1 were mixed to form 'serious crime set 1' and the same experimental image set was mixed with decoy set 2 to form 'volume crime set 2'. The position of the decoy marks was randomly assigned using a random number generator for each set.



The order of the set of experimental marks was randomised once and these marks were then placed around the position of decoy marks in both photo sets. This resulted in two sets of images containing the same experimental marks in the same random order, but with a different position and set of decoy marks in each. Order and content information for both sets can be found in Figure 3.2.

Figure 3.2 - Formation of experimental image sets



Both sets of mark photographs were bound in hardback A6 notebooks, with one image per page. Each photograph was manually cropped to ensure that only the ridge detail of interest was visible and that previous annotations were removed. The reference letter and the crime type to which the mark related was written above each mark on the top right hand corner of the page. The serious crime photograph set was accompanied with a 'serious crime' instruction sheet (Appendix B.5), and results sheet (Appendix B.6), whilst the volume crime photograph set was accompanied with a 'volume crime' instruction sheet (Appendix B.7) and results sheet (Appendix B.8). Instruction sheets outlined how to complete the experimental task, and only differed between the two crime contexts in that the 'volume crime' instruction sheet stated that practitioners should "Decide whether or not you would mark this fingerprint up **in a case of Theft from Vehicle**", whereas the 'serious crime' instruction sheet stated

“Decide whether or not you would mark this fingerprint up **in a case of Murder**”. Results sheets also specified the crime type to be either ‘Theft from Vehicle’ or ‘Murder’. No further contextual information was provided to practitioners. Both sets of mark photographs were also accompanied by an information sheet (Appendix B.9) and consent form (Appendix B.10). Justification for all aspects of material production is presented in Table 3.1.

Table 3.1 - Justification of aspects of material production

Stage of Method	Aspect of Material Production	Justification
Mark development	Fingermarks were deposited on ‘clean’ white sheets of paper	This ensured that all developed marks were likely to have been of known source for later attempted identification, and prevented background interference
	Marks were deposited with a range of pressures and movements	Different pressures and movements were applied during mark deposition at random in order to provide a range of quality and appearance in developed marks
	A combination of natural and amino acid pad deposition was used	To ensure that marks were of mixed quality and that the development of some marks would occur
	Ninhydrin treatment was used as the development method for experimental marks	Ninhydrin often produces broken up ridge detail which can make quality difficult to determine making it preferable for developing a set of challenging fingerprints. In addition, treatment with ninhydrin is a one stage process requiring less participation time from operational stakeholders
	Clearly good and poor quality marks were removed from the image set prior to examiner assessment	This ensured that the time of fingerprint examiner participants was spent considering the more challenging marks developed. This process was carried out by the researcher (who had experience as a laboratory practitioner) and an operational laboratory practitioner
Examiner assessment of marks	Three sets of tenprints were taken including all areas of ridge detail	To ensure that all areas of ridge detail were captured to the best possible quality for identification purposes
	Tenprints were recorded under an assumed name	To increase the ecological validity of the study whilst protecting the anonymity of the researcher
	Examiners measured comparability of marks as through attempting to identify them to the tenprints	This gave a more objective and ecologically valid judgement of the suitability of the mark for comparison
Assembly of experimental image set	20 experimental marks were selected for inclusion in the experimental image set	This enabled a sufficient number of decisions to be made throughout the experiment (based upon similar sample sizes utilised elsewhere e.g. Dror et al. 2011) whilst limiting the requirements for participation time within the operational laboratories
	The opinions of the five examiners were consistent on the comparability of the 20 experimental marks selected (with 1 exception)	This allowed the research to be based on marks which had been deemed comparable or insufficient by more than one examiner and removed any marks from the experimental set that had resulted in variability of opinion

The experimental mark set contained marks of a mixture of qualities	Marks were primarily considered to be borderline for submission but were a mixture of better and poorer quality marks so as to ensure a range of participant decisions
Different decoy marks were included in each of the image sets	To give the appearance that the two sets of marks were different
Examiner opinion did not need to be in agreement for decoy marks	Participant responses to these marks were not analysed
The decoy marks were ordered differently in both image sets while the order of the experimental marks remained the same	To give the appearance of two different sets of images without causing order effects in the experimental images. The first two images of each set were decoy marks to suggest that all marks to follow would be different.
Images were bound in hard back books, one per page	To prevent marks being assessed simultaneously or in a different order to that presented
Images were manually cropped to remove all extraneous information, except the context and mark reference	To remove any additional extraneous information but to ensure that the crime context of the mark was considered in each case
Fingerprint images were used in the study as opposed to original marks	This prevented the possibility that the quality of the mark used in the study would change over time (fingermarks developed with ninhydrin can continue to develop over time), and therefore may differ between the occasions of examiner and practitioner assessment. Additionally, the use of mark photographs allowed multiple copies of each mark to be made so that practitioners could participate in the study simultaneously, increasing the rate of participation possible

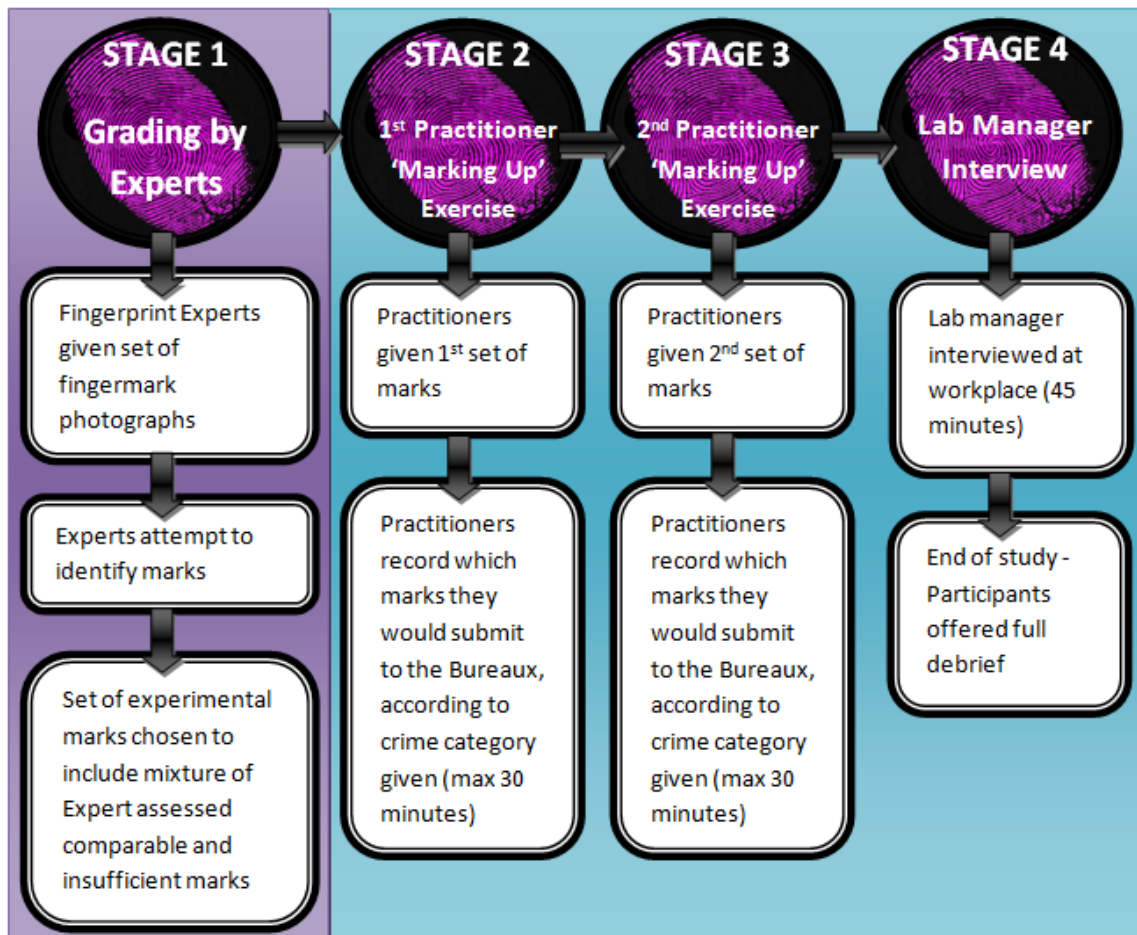
### 3.3.2.2. Participants

Participants were the 13 fingerprint practitioners working within a large metropolitan UK Fingerprint Visualisation Laboratory (Laboratory A). This constituted all available practitioners working within the laboratory (with the exception of those on long term leave or seconded to an external role). They had a mean value of 8.95 years of experience, with a minimum experience of 10 months and a maximum of 18 years. All participants had received initial training at the College of Policing, and had received continued training internally.

### 3.3.2.3. Procedure

An overview of the method of the study, including generation of test marks is provided in Figure 3.3. Please note, timings refer to approximate participation time required, and were not communicated to participants as a time constraint.

Figure 3.3 - An overview of the experimental method



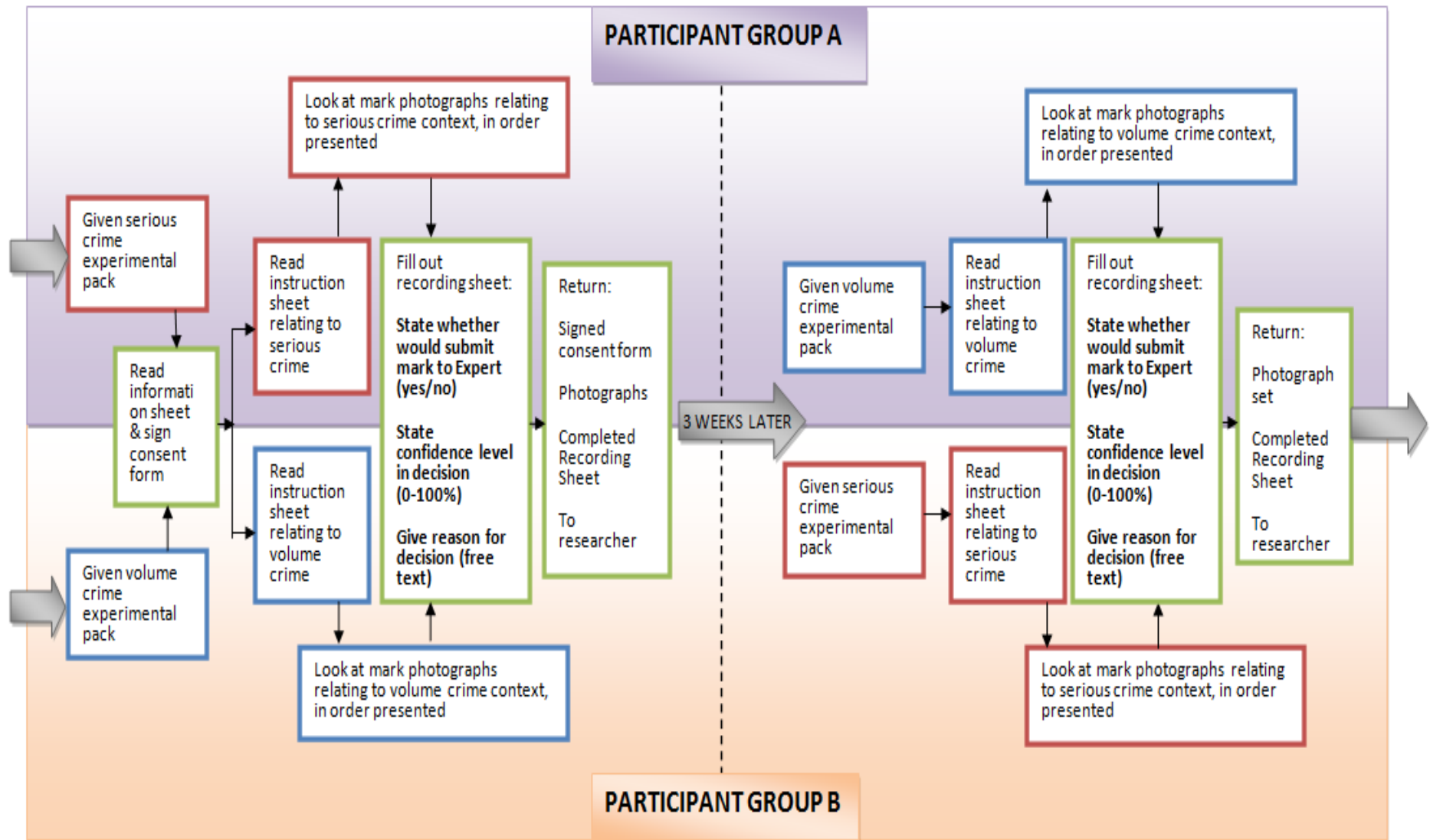
Participants were asked to complete the experimental task on two occasions. On each occasion, each participant was given either the volume or a serious crime set of experimental fingerprint photographs. A gap of at least three weeks was left between the two experimental tasks as this was deemed to be a sufficient amount of time to ensure that the experimental fingerprints were not recognised, particularly given that practitioners were likely to have assessed many additional marks within casework during this time. Participants were asked to complete the response sheet provided alongside the image set stating, for each mark, in the order presented, whether they would submit the mark to a fingerprint examiner, how confident they felt in this decision, and to state the reasons for making this decision. Submission decisions were recorded as either a yes or no response, preventing participants from providing less certain responses. Participants were asked to rate their confidence on a full confidence scale (0-100%). This allowed participants to give an entirely subjective opinion of their confidence, as opposed to a more restricted opinion that would have been possible with the use of a half range scale (50-100%) which is often used in confidence assessment of tasks with two possible 'answers' (Adams & Adams, 1961). A within-subjects design was used in order to counterbalance the study in relation to crime context. Participants were, therefore, randomly

assigned to two groups: A (N=7) and B (N=6), to form a within-subjects design in relation to crime context. Group A completed the marking up task in relation to the serious category crime on the first occasion, whilst Group B completed the task in relation to a volume crime first. Participants were instructed to return the photo book and completed results sheet after completing each task in order to prevent any direct comparison of the contents of the two books and a three-week gap was left between participation in the two tasks in order to avoid the recognition of repeated fingerprint images within the two image sets. Participants were asked to carry out the tasks at their normal workplace, under normal working conditions, using standard practices and equipment in order to maximise the ecological validity of the results. The experimental task was self-completed and the laboratory manager distributed the tasks and ensured that participants were working in the same environment in which they would quality assess casework marks with access to the same equipment (eye glasses and appropriate lighting conditions). Participants were instructed to spend the same amount of time considering each mark as they would during standard casework. This instruction was intended to avoid a floor or ceiling effect in time taken. The procedure followed is outlined in Figure 3.4.

#### 3.3.2.1. Laboratory manager policy interview

The laboratory manager (also one of the 13 practitioner participants) was interviewed after completing the second fingerprint assessment task. The interview was semi structured allowing inclusion and pursuit of interesting areas of discussion raised, whilst providing a framework to ensure all desired information was gained. The interview took place in an empty meeting room within the laboratory premises and lasted approximately 45 minutes. Questioning took place under the areas of business structure, recruitment and training, fingerprint recovery and submission policy and quality assurance practices. The interview schedule is provided in Appendix B.11. The interview exploited the 'insider-outsider' status of the researcher (Corbin Dwyer & Buckle, 2009), who had previously worked in a fingerprint laboratory, through the use of relevant terminology and empathy for the challenges of the role, which helped to build trust during the interview. The interview was tape recorded and transcribed.

Figure 3.4 - Experimental methodology



### 3.3.3. Method for objective 3.4

The methodology outlined in relation to Objectives 3.1-3.3 had previously been utilised within the Metropolitan Police Service Serious Crime Fingerprint Evidence Recovery Unit (Earwaker et al. 2015). Within this original study the 11 available laboratory practitioners within the Serious Crime Fingerprint Enhancement Laboratory (all practitioners employed within the laboratory who were not on leave at the time of participation) were asked to carry out the experimental task as stated in the experimental design of Objectives 3.1-3.3 detailed within 3.3.1. As per the method stated in 3.3.2.3. participation in the study was self-completed in the normal environment used within the laboratory for quality assessing fingermarks, with access to the same equipment. The purpose of Objective 4 was to compare the original data previously published by Earwaker et al. (2015) with that generated by the application of the research methodology to Laboratory A. As such, data analysed in relation to Objective 4 was gained through the application of the method provided in 3.3.1 to both The Metropolitan Police Service Laboratory (referred to as MPS) and Laboratory A. Each force participation in the experiment by employing their standard working conditions to the experimental task, this allowed a comparison of results according to standard processes, increasing the ecological validity of the task.

## 3.4. Results

### 3.4.1. Data grouping

Categorising laboratory A data within Objectives 3.1 - 3.3

Each decision made by participants was categorised according to whether it was in agreement with the decision of the fingerprint examiners (i.e. the practitioner had submitted a fingermark that was comparable or had discarded a mark that was not comparable), a false-positive result (the practitioner had submitted a fingermark that was not comparable), or a false-negative result (the practitioner had discarded a fingermark that was comparable). For the purposes of some analyses false-positive and false-negative results were group as 'erroneous' in that they went against the inferred suitability conclusions of the examiners (the ground truth). The use of an examiner usability determination as a 'ground truth' for data analysis is later discussed within 7.5.1.

Data sets for comparison within Objective 3.4

The categorised data analysed within Objectives 3.1-3.3 was compared with the findings of preliminary research published by Earwaker et al. (2015) in order to compare findings in relation to mark submission efficiency, effect of crime context, and self-reported confidence between Laboratory A and the Metropolitan Police Serious Crime Fingerprint Laboratory.



Data files relating to Objectives 3.1- 3.4 are provided in Appendix B.12.

### 3.4.2. Objective 3.1: Assessment of the efficiency of fingerprint submission decision-making within Laboratory A

Data analysis under this objective aimed to establish the extent to which fingerprint laboratory fingerprint submission decisions matched the usability determinations of fingerprint examiners.

13 participants each made 2 submission decisions on each of the 20 experimental mark photographs, giving an overall total of 520 submission decisions. This includes fingerprints presented in both crime categories so as to gain an overall picture of the efficiency of fingerprint submission.

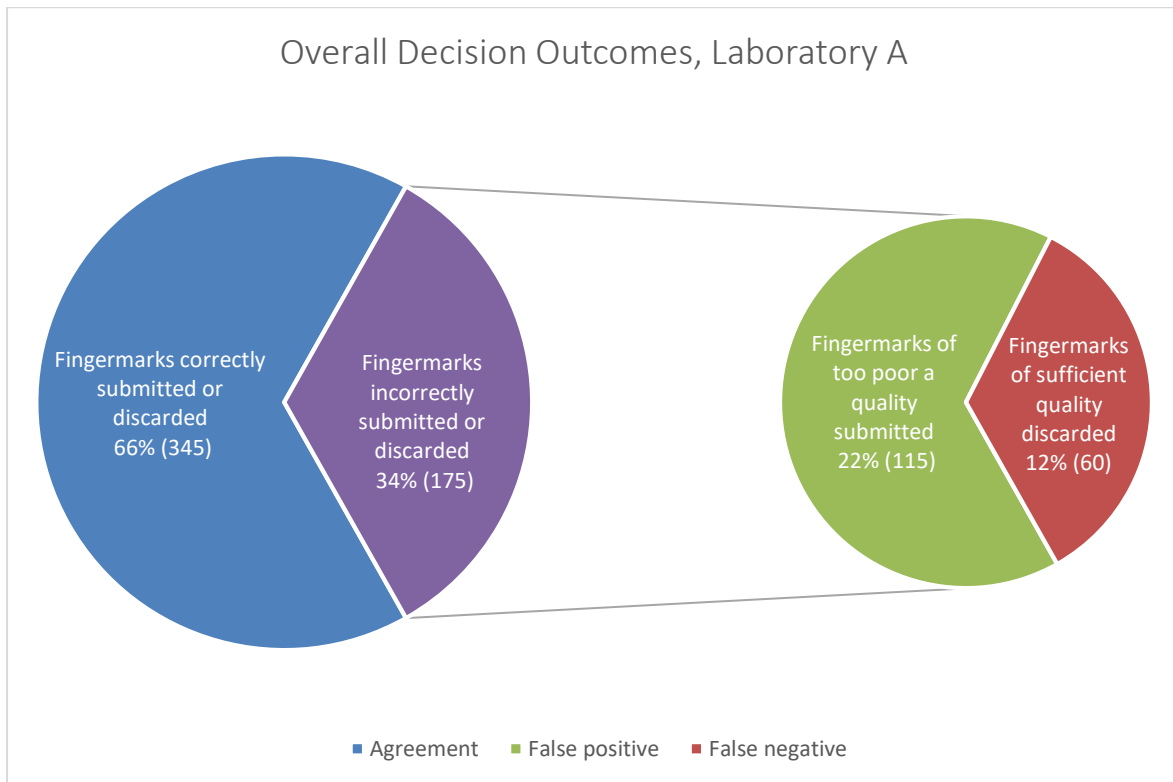
#### Overall decision efficiency

When considering all fingerprint submission decisions made, 66.35% of decisions were correct in relation to the usability decisions reached by the fingerprint examiners. 33.65% of decisions made were erroneous; not matching the usability consensus reached by the fingerprint examiners. 65.71 % of the erroneous decisions were false positives (fingerprints of too poor a quality submitted), and the remaining 34.29% of erroneous decisions were false negatives (fingerprints of sufficient quality not submitted). Of all decisions made 22.12% (115 decisions) were false positives and 11.54% (60 decisions) were false negatives.

Figure 3.5 shows the overall percentage breakdown of decision outcomes. False-positive decisions can be seen to be responsible for almost two thirds of the erroneous decisions, with false negative decisions making up the remaining approximate third of erroneous decisions. This means that the hypothesis can be accepted that the practitioner's decision threshold can be too high, resulting in false-negative results, but that the more common position is that that submission threshold is too low as false positive results make up two thirds of the erroneous decisions.



Figure 3.5 - Pie chart of overall decision outcomes within Laboratory A



Decision outcomes according to examiner assessment ('ground truth')

Figure 3.6 shows the percentage of 'erroneous' decisions as calculated according to examiner assessment of the usability of the fingerprint. A higher percentage of erroneous decisions were made in relation to marks examiner assessed as insufficient than were made in relation to marks assessed as sufficient. A Chi Squared test showed that the difference between the proportion of erroneous and decisions in agreement for examiner-deemed sufficient and insufficient fingerprints was significant ( $\chi^2, 1) = 26.358, p < 0.001$ . The Phi test statistic was 0.225 which indicates that the relationship between a correct or erroneous decision outcome and the examiner deemed usability of the fingerprint was, however, weak. The contingency table for this analysis is provided in Table 3.2

Table 3.2 - Contingency table for Chi Squared test for difference in proportion of erroneous and agreement outcomes for sufficient and insufficient fingerprints

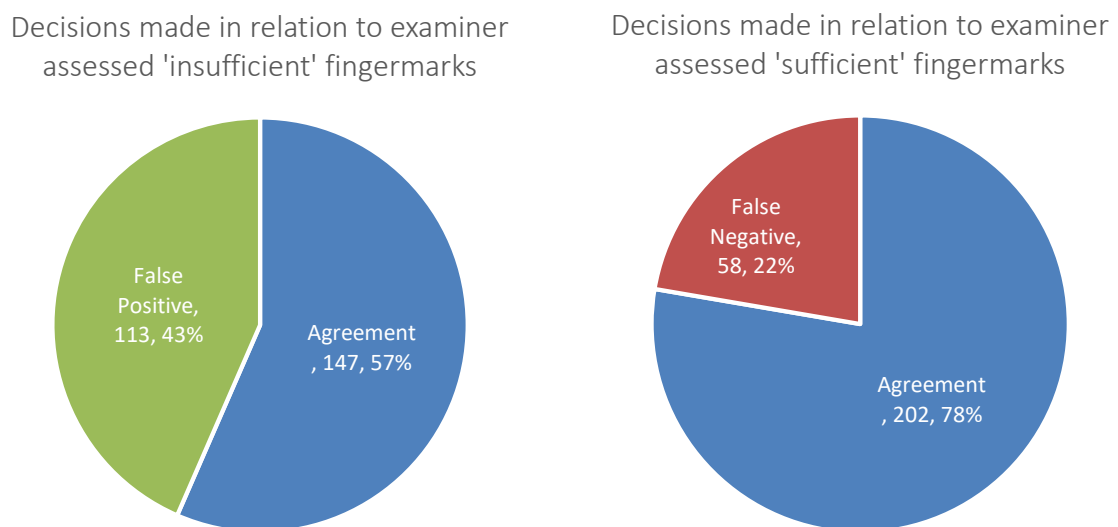
			Mark Type		Total
			Insufficient	Sufficient	
Outcome	Agreement	Count	147	202	349
		Expected Count	174.5	174.5	349
	Erroneous	Count	113	58	171
		Expected Count	85.5	85.5	171

Total	Count	260	260	520
	Expected Count	260	260	520

Erroneous decisions in insufficient fingerprints and correct decisions in sufficient fingerprints were overrepresented, while correct decisions in sufficient fingerprints and erroneous decisions in insufficient fingerprints were underrepresented. Erroneous decisions in the case of insufficient marks were the most overrepresented decision type, while erroneous decisions in the case of sufficient marks were the most underrepresented decision type. Thus, practitioners demonstrated greater accuracy in their decisions in relation to sufficient fingerprints.

The use of an inferential statistical test here is potentially problematic as multiple observations by the same participant can violate the assumption of independent observations. Here observations in relation to the same fingerprint were made on two occasions by the same practitioner. In order to evaluate the statistical results presented, decision outcomes were also examined at an individual practitioner level to assess consistency in performance across participants (see 'Decision outcomes according to practitioner').

Figure 3.6 - Decision outcomes according to examiner mark assessment



#### Error Rates, Sensitivity, and Specificity

Drawing upon analysis undertaken by Langenburg et al. (2012) a number of calculations were carried out in relation to the examiner determination (or the 'ground truth') of the experimental fingerprints.

According to Langenburg et al. (2012) an error rate can be calculated to establish the proportion of cases in which a submission decision was made, given that the fingerprint was insufficient (a false

positive error rate), and the proportion of cases in which a discard decision was made, given that the fingerprint was sufficient (a false negative error rate) (Langenburg et al., *ibid*). Sensitivity (the proportion of reported submissions when marks could be identified) and Specificity (the proportion of discard decisions when the images were insufficient) were also calculated. This data is presented in Table 3.3.

Table 3.3 - Tabulated error rates, sensitivity, and specificity

False Positive Error Rate	False Negative Error Rate	Sensitivity	Specificity
43.46%	22.31%	77.69%	56.54%

Decision outcomes according to marks forwarded or discarded

In relation to the decision made by practitioners to submit or discard a fingerprint, the proportion of cases in which the fingerprint was insufficient given a submission decision has been reported, and the proportion of cases in which the fingerprint was sufficient given a discard decision has been reported can be calculated. Langenburg et al. (2012) refer to these calculations as a false positive discovery rate and false negative discovery rate, respectively. These figures are provided in Table 3.4.

Table 3.4 - Tabulated false positive and false negative discovery rates

False Positive Discovery Rate	False Negative Discovery Rate
35.87%	28.29%

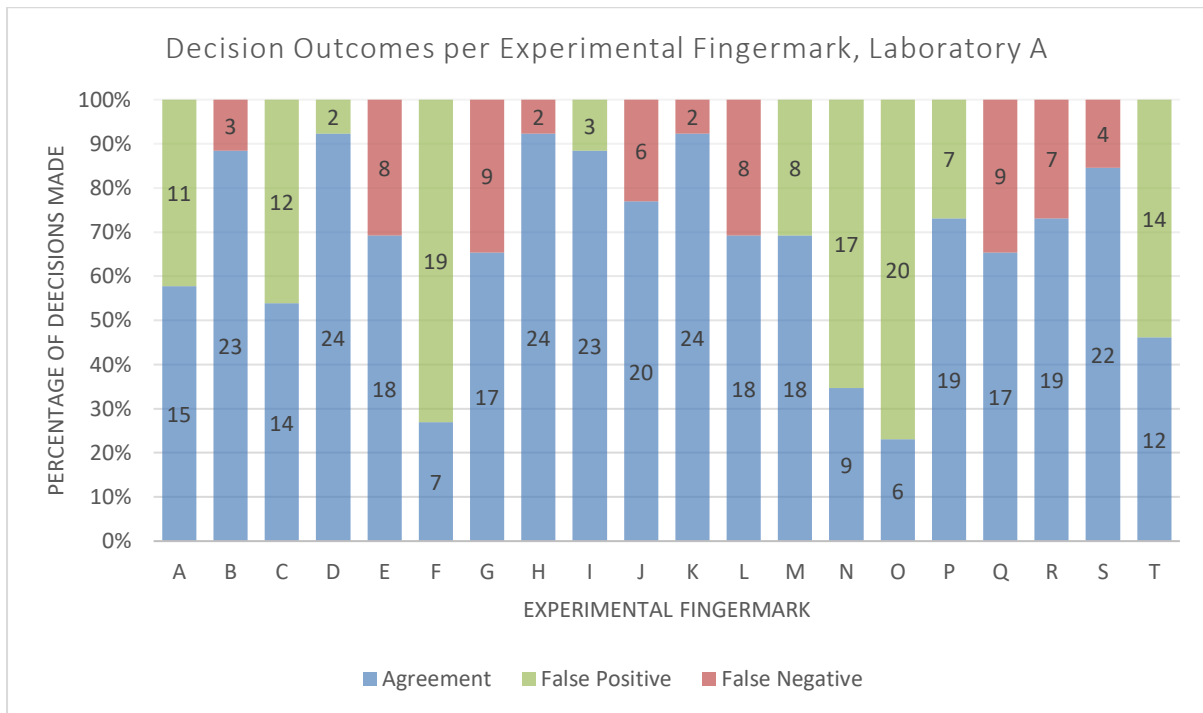
The presentation of the values in Table 3.3 and Table 3.4 allows consideration of the level of error within the fingerprint submission process (in comparison to examiner ground truth usability) in standard terminology used elsewhere with fingerprint decision making performance literature (Langenburg et al., 2012). Summarising the data in this format may be of benefit for future comparison with similar data within fingerprint decision making as it provides a standard approach to reporting findings, including those relating to different aspects of the fingerprinting process, so as to enable future comparison of efficiency across stages of the process.

Decision outcomes according to experimental fingerprint

For each of the 20 fingerprints a total of 26 decisions were made, 2 by each of the 13 practitioners within the laboratory.

The percentage and numerical make up of decisions outcomes for each experimental mark is represented in Figure 3.7.

Figure 3.7 - Decision outcomes according to experimental fingermark



Divided according to experimental fingermark, the data shows that there is not a consistent proportion of error in relation to each mark, rather that the proportion of error varies between marks and there are particular marks for which there was a considerably higher level of erroneous decisions made. In particular fingermarks F, N, O, and T can all be seen to have a percentage of erroneous decisions higher than 50%.

Decision outcomes according to practitioner

Each of the 13 practitioners made a total of 40 decisions, 2 concerning each experimental fingermark. The combined outcomes of these decisions for each practitioner are presented in Figure 3.8. Table 3.5. provides a summary of descriptive statistics in relation to decision outcomes across the participating practitioners.

Figure 3.8 - Decision outcomes according to practitioner

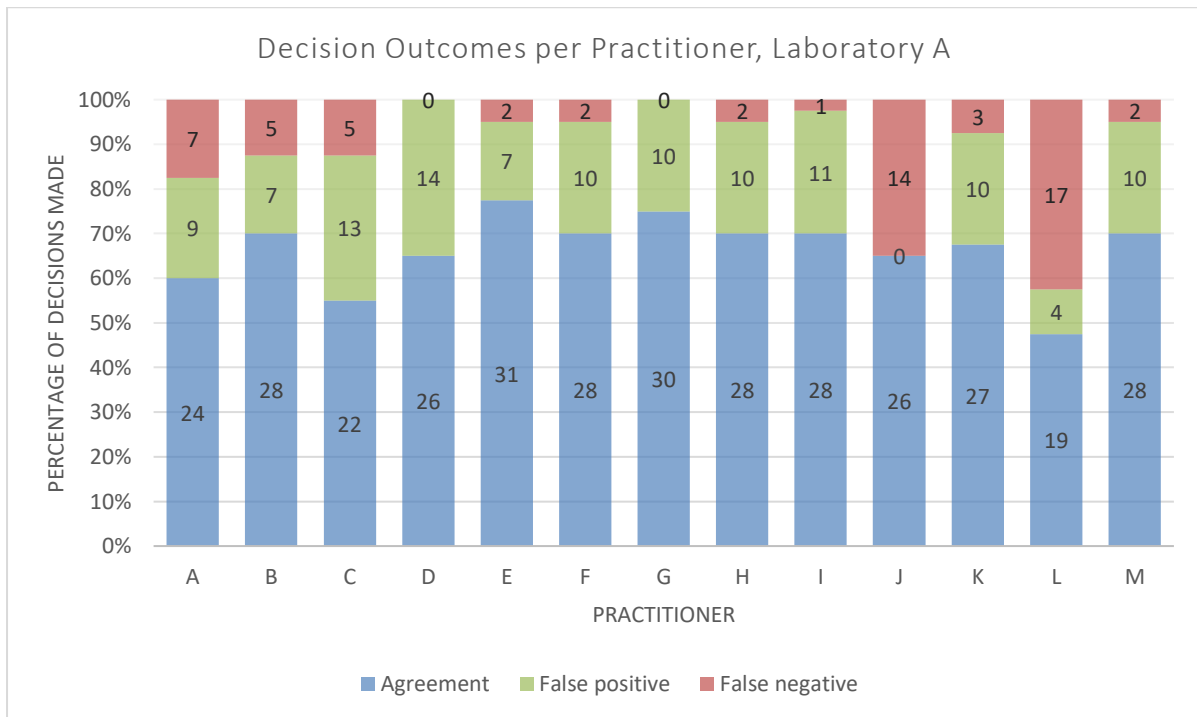


Table 3.5 - Descriptive statistics in relation to decisions made by all practitioners

	Agreement with Ground Truth	False Positive Decision	False Negative Decision
Mean	26.54	8.85	4.62
Variance	10.60	13.64	27.76
Range	12	14	17
Standard Deviation	3.26	3.69	5.27

Dividing the data according to practitioner provides insight into the variation in decision outcomes across the laboratory. Figure 3.8 demonstrates that erroneous decisions were made by all practitioners during the study, rather than error being confined to a small number of outlying practitioners. The data shows a common pattern in decision outcomes. Practitioners made correct decisions in the majority of cases, with a moderate percentage of false positive errors and a small percentage of false negative errors. There is some variation in decision outcomes between practitioners and also some exceptions to the general trend. Two practitioners (J and L) have particularly high percentages of false negative decision outcomes, while two practitioners (D and G) did not make any false negative decisions. As a result of these outliers the range, standard deviation and variance of false negative decisions is higher than that of decisions in agreement with examiners and false positive decisions (see Table 3.5).

## Key findings from laboratory manager interview

The semi-structured interview with the laboratory manager of Laboratory A (who was also one of the participating practitioners) resulted in the following key information in relation to the procedure for fingerprint submission from the fingerprint laboratory to the fingerprint bureau:

- There was no procedure within the laboratory for determining which fingerprints should be marked up and submitted to the bureau
- A feedback mechanism for the detection of excessive false positive mark submissions was in place in the form of examiners reporting this position, but there was no mechanism for the routine detection of false negative erroneous decisions, as per the opinion of a fingerprint examiner
- Some interaction between the bureau and laboratory did take place in relation to the mark submission process. In the case of some evidence relating to serious crime, examiners would come to the laboratory to select which fingerprints marked up by the practitioner they wanted to be submitted to the bureau. There was no documented criteria or procedure for this examiner quality assessment.
- The laboratory operated within a 'take the best' strategy, meaning that, in volume crime cases, practitioners would mark up only 'the best' 20 marks visualised upon an exhibit. No procedure, however, was outlined for defining comparative quality between fingerprints
- There was no different policy or procedure in relation to individual mark quality assessment according to crime context
- Training for laboratory practitioners was carried out in-house as part of an induction and mentoring process. This did involve some input from the fingerprint bureau

These key results in relation to the fingerprint submission process indicated that there was no existing criteria governing which marks should be submitted. Equally there was no prescribed difference in the quality or number of marks that should be submitted according to crime context. This key information validates the findings of data analysis within this chapter as there was not found to be policy or procedure to indicate that practitioners should have been submitting marks in a certain way and, thus, illustrates that policy or procedure cannot be provided as an explanation for the fingerprints submitted and discarded in this study

## Summary of Objective 3.1. results

Objective 3.1. set out to establish the efficiency of fingerprint submission within Laboratory A through comparing the submission decisions of fingerprint laboratory practitioners with the usability determinations of fingerprint examiners. Analysis of the data has shown inconsistencies between the

submission decisions made and the usability of the mark by a fingerprint examiner. Erroneous decisions were made in relation to all of the experimental fingerprints considered and by each of the participating practitioners, demonstrating that this overall effect was not due to a minority of fingerprints or practitioners.

3.4.3. Objective 3.2: The effect of crime context on fingerprint submission decision-making within Laboratory A

Date analysis under this objective tested the hypothesis that case contextual information would affect the fingerprint submission decision making thresholds of practitioners, resulting in lower thresholds in the case of serious crime contextual information and higher thresholds in the case of volume crime contextual information (as was reported by Earwaker et al (2015)).

Assessing for order effects within the data

The counter-balanced nature of the study meant that it was important to establish if there was a significant difference between the decision outcomes for the ‘time 1’ and ‘time 2’ data sets. A statistical difference may suggest that there had been carry-over effects (for example tiring of the tasks and so performing less well on the second occasion) or practice effects (enhanced performance on the second occasion due to practice), detracting from any observed contextual effects.

The results of a Chi Squared test showed no significant difference between expected and observed decision outcomes according to the time set in which the fingerprints were presented (N= 520, p=0.860), suggesting the absence of carry-over or practice effects. The contingency table for this analysis is presented in Table 3.6.

Table 3.6 - Contingency table for Chi Square test for difference in expected and observed decision outcomes according to time of presentation

			Time		Total
			Time 1	Time 2	
Outcome	Agreement	Count	171	174	345
		Expected Count	172.5	172.5	345
	False Neg	Count	32	28	60
		Expected Count	30	30	60
	False Pos	Count	57	58	115
		Expected Count	57.5	57.5	115
Total		Count	260	260	520
		Expected Count	260	260	520

On both occasions (Time 1 and Time 1) the majority of decisions were in agreement with the usability determinations made by the fingerprint examiners (approximately 65.77% and 66.92% respectively), with submitting poor quality marks accounting for approximately 21.92 and 22.31 % of decision outcomes, and discarding good quality marks accounting for between approximately 12.31 and 10.77 % of decision outcomes respectively.

Overall decision profile according to crime category

Data was divided according to the crime category in which the fingerprints were presented, resulting in 260 decisions per crime category, made up of 13 participants making one decision on each of the 20 fingerprints.

A Chi – Square test showed that the relationship between crime context and decision outcome was not statistically significant ( $\chi^2, 2) = 1.043, p = 0.593$ . The contingency table for this analysis is provided in Table 3.7.

Table 3.7 - Contingency table for Chi Square test for relationship between crime context and decision outcome

			Crime Context		Total
			Serious	Volume	
Outcome	Agreement	Count	170	175	345
		Expected Count	172.5	172.5	345
	False Neg	Count	28	32	60
		Expected Count	30	30	60
	False Pos	Count	62	53	115
		Expected Count	57.5	57.5	115
Total		Count	260	260	520
		Expected Count	260	260	520

In relation to both crime categories the majority of decisions were in agreement with the usability determinations made by the fingerprint examiners (66.38% in the serious crime category, and 67.31% in the volume crime category). Submitting poor quality marks accounted for approximately 22.85% of decision outcomes in the serious crime context, and 20.38% of decision outcomes in the volume crime context. Discarding good quality marks accounted for 10.77% of decision outcomes in the serious crime context, and 12.31% in the volume crime context.

Whilst there was no statistically significant difference in decision outcomes according to case context, there can be seen to be a slight variation in decision outcomes. In volume context there is a slightly



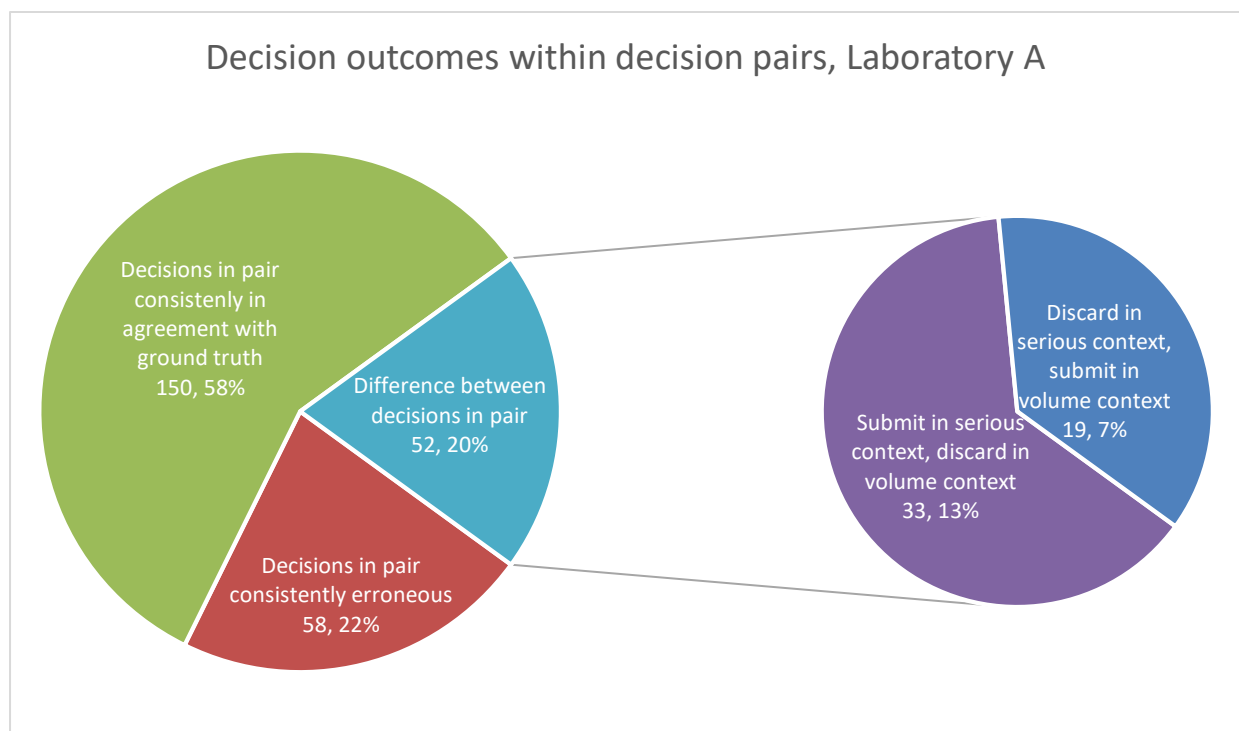
higher percentage of correct decision outcomes, a slightly lower proportion of false positive decisions, and a slightly higher proportion of false negative decisions. This may suggest the adoption of a slightly lower, but not statistically significant, submission threshold in the context of a serious crime.

### Decision Consistency

Data were interrogated to determine the level of consistency in the submission decisions made by practitioners. This was possible as each practitioner had made two decisions in relation to each experimental fingerprint during the course of the experiment. This analysis was performed in order to further investigate where similarities according to crime context were located within the data set, and to establish to what extent the similarities in decisions outcomes according to crime context were due to practitioners making the same decisions in relation to the same fingerprints in the two contexts or whether practitioners were making different decisions in each context, but that the differences in these decisions were being counteracted by opposite decision making by other practitioners in the data set.

In total there were two decisions made by each of the 13 practitioners about each of the 20 experimental fingerprints, resulting in 260 pairs of decisions (two decisions made by each practitioner in relation to each mark on the two occasions). Each pair of decisions was analysed to investigate consistency in decision making. This data is presented graphically in Figure 3.9.

Figure 3.9 - Pie chart illustrating decision consistency



208 of the 260 pairs of decisions were consistent (the practitioners made the same decision in relation to the same mark on both occasions on which it was presented). 150 of these consistent decisions were in agreement with the established ground truth and 58 of these consistent pairs were erroneous in relation to the usability determinations of the fingerprint examiners.

Of the 260 decision pairs there were 52 occasions in which the decision of a practitioner changed when viewing the same fingermark on the first and second occasion; in 20.00% of the decision pairs the decision was inconsistent.

Inconsistencies in decision pairs were further broken down to establish in which context and which presentation time the decision was to submit, and in which it was to discard the fingermark. In 33 of the pairs of inconsistent decisions practitioners had submitted a fingermark when it was presented in a serious context, but had not submitted the same fingermark when it was presented in a volume context. This accounted for 63.46% of the total pairs of decisions that changed between time 1 and time 2, and 12.69% of all the decision pairs. This type of decision change occurred a similar number of times when the serious context was presented on time one and the volume context on time two, and when the volume context was presented on time one and the serious on time two (in 16 of these 32 cases the serious context was presented first and in 17 of these cases the volume context was presented first). In 19 of the pairs of inconsistent decisions practitioners had submitted a fingermark when it was presented in a volume context, but had not submitted the same fingermark when it was presented in a serious context. This accounted for 36.54% of the total pairs of decisions that changed between time 1 and time 2, and 7.31% of all the decision pairs. This type of decision change occurred 12 times when the serious context was given at time 1, and 7 times when the serious context was provided at time 2.

Decision consistency according to experimental fingermark

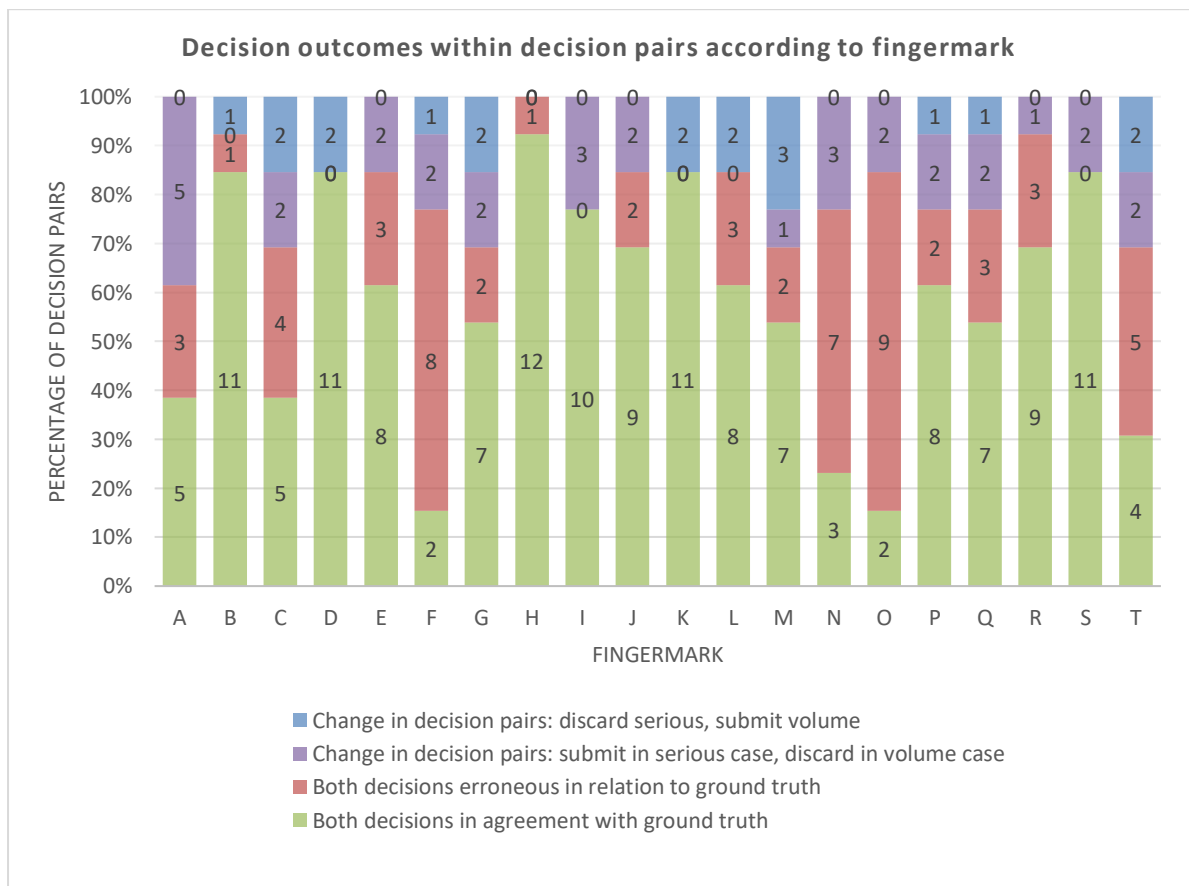
Data in relation to decision consistency was divided according to experimental fingermark. The data presented in Figure 3.10 represents the number and percentage of occasions in which:

- The decision made by a practitioner in relation to the fingermark was in agreement with the 'ground truth' on both occasions
- The decision made by a practitioner in relation to the fingermark was erroneous on both occasions
- The decision made by a practitioner in relation to the fingermark changed; when the mark was presented in a volume context the practitioner decided to discard the fingermark, but when presented in a serious context the practitioner submitted the same mark

- The decision made by a practitioner in relation to the fingerprint changed: when the mark was presented in a volume context the practitioner decided to submit the fingerprint, but when presented in a volume context the practitioner decided to discard the same mark

Variation in decision consistency and the direction of changed decisions can be seen between fingerprints. For all marks the majority of practitioners were consistent in their decision making. For the majority of marks most consistent pairs of decisions were 'correct', but for some marks the predominant occurrence was that both decisions were erroneous, highlighting the variation between marks and suggesting a distinction between more straightforward and more challenging marks. Inconsistent decisions varied between 7.69% and 38.46% of decision pairs. Where there have been inconsistent decisions made in relation to a mark these have predominantly been either all to submit in the serious context and discard in the volume context or mixed between submitting in the serious case and discarding in the volume case, and discarding in the serious case and submitting in the volume case. There are only three fingerprints for which all inconsistent decision pairs were to submit in the volume case and discard in a serious case (B, D and K) and these had very low levels of inconsistency.

Figure 3.10 - Decision consistency according to experimental fingerprint

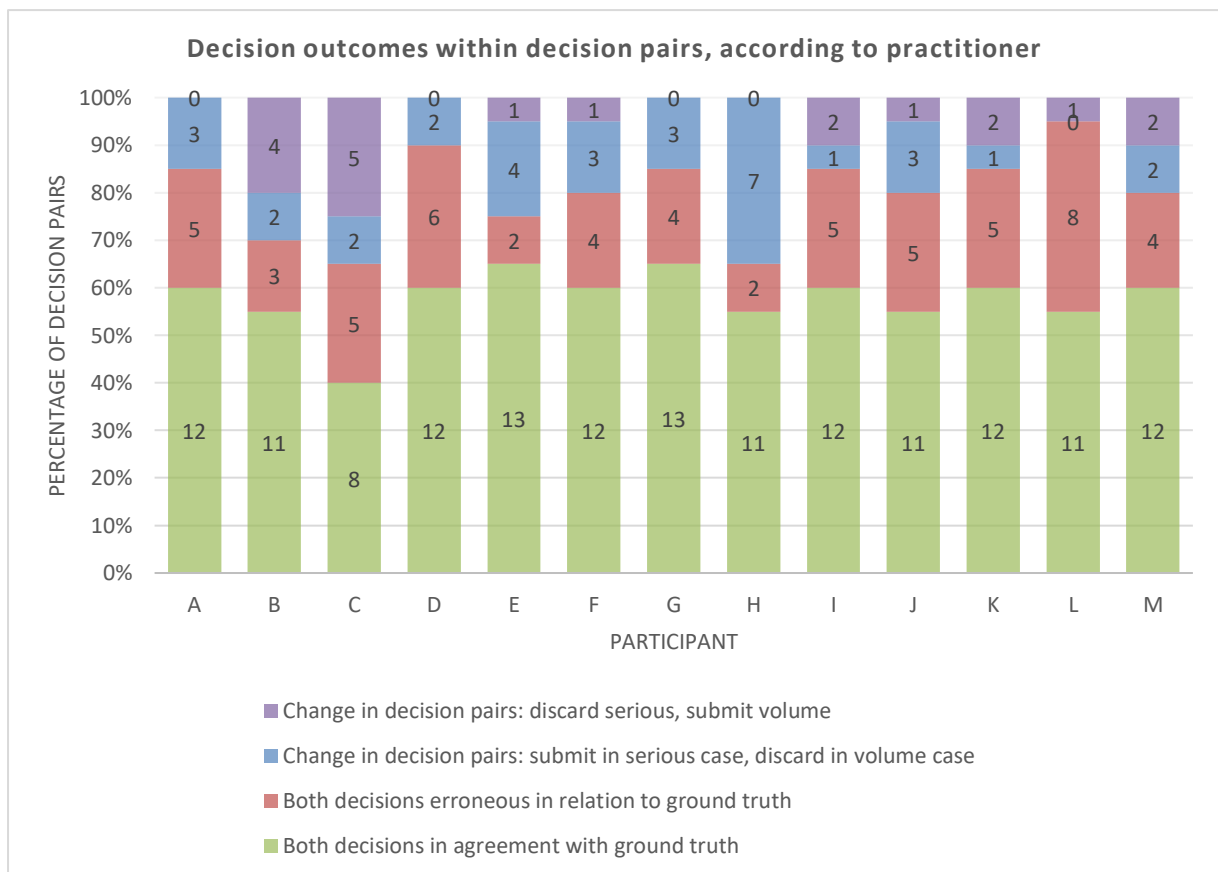


### Decision consistency according to practitioner

Data was divided to illustrate decision consistency for each practitioner, with each practitioner making 20 decision pairs (one pair of decisions in relation to each of the 20 experimental fingermarks). Decision pairs were categorised in the same four categories as previously described. Participant decision consistency data is presented in Figure 3.11.

All but one of the practitioners (C) made correct decisions in both contexts in relation to the majority of cases. All practitioners made two erroneous decisions in relation to a smaller proportion of marks. The majority of practitioners made changed decisions to submit marks in serious cases and discard in volume cases as well as to discard in the serious context and submit in the volume context. Three practitioners (A, D, and H) only submitted in the serious context and discarded in the volume context when there were inconsistencies in their decision making in relation to the same marks, and one practitioner (L) only submitted in the volume context and discarded in the serious context when there was an inconsistency in their decision making, although this only happened on one occasion.

Figure 3.11 - Decision consistency according to practitioner



### Summary of Objective 3.2. results

Objective 3.2. set out to establish the effect of contextual information in the form of crime type on the submission decisions made by laboratory practitioners. Analysis has shown there to be no significant difference in mark submission according to the type of crime in which fingermarks were presented ( $\chi^2, 2 = 1.043, p = 0.593$ ), in contrast to the hypothesis suggested based upon the findings of Earwaker et al. (2015). Examining the effect of crime context at an individual decision level illustrated that in cases in which practitioners had made different decisions in relation to the same mark they had changed their submission decision to submit in a serious case and discard in a volume case more often than changing from a decision to submit in a volume case to discard in a serious case. All practitioners were found to change their decision in relation to the same mark at least once, and practitioner decisions changed in relation to all but one fingermark.

### 3.4.4. Objective 3.3: Self-reported confidence

Data analysis under this objective set out to test the hypothesis that laboratory practitioners will be overconfident in their fingermark submission decision making, through analysing self-reported confidence ratings in the decisions made as part of the mark submission exercise.

#### Comparison of self-reported confidence levels

The confidence levels reported by practitioners in relation to each of the 520 decisions made were collated to provide an overall picture of submission decision confidence, and were also divided according to the category of decision that they related to so as to allow for an investigation of the experimental factors that affected this self-reported confidence.

Categories of decisions for the purposes of investigating their effect on confidence levels were:

- Time submission decision was made (first presentation of fingermark (T1) or second presentation of fingermark (T2))
- Context of submission decision made (volume crime or serious crime)
- Type of decision made (to submit or to discard fingermark (irrespective of ground truth))
- Outcome of decision made (outcome in agreement with ground truth or erroneous in relation to ground truth)
- Type of erroneous decision made (false positive decision or false negative decision)

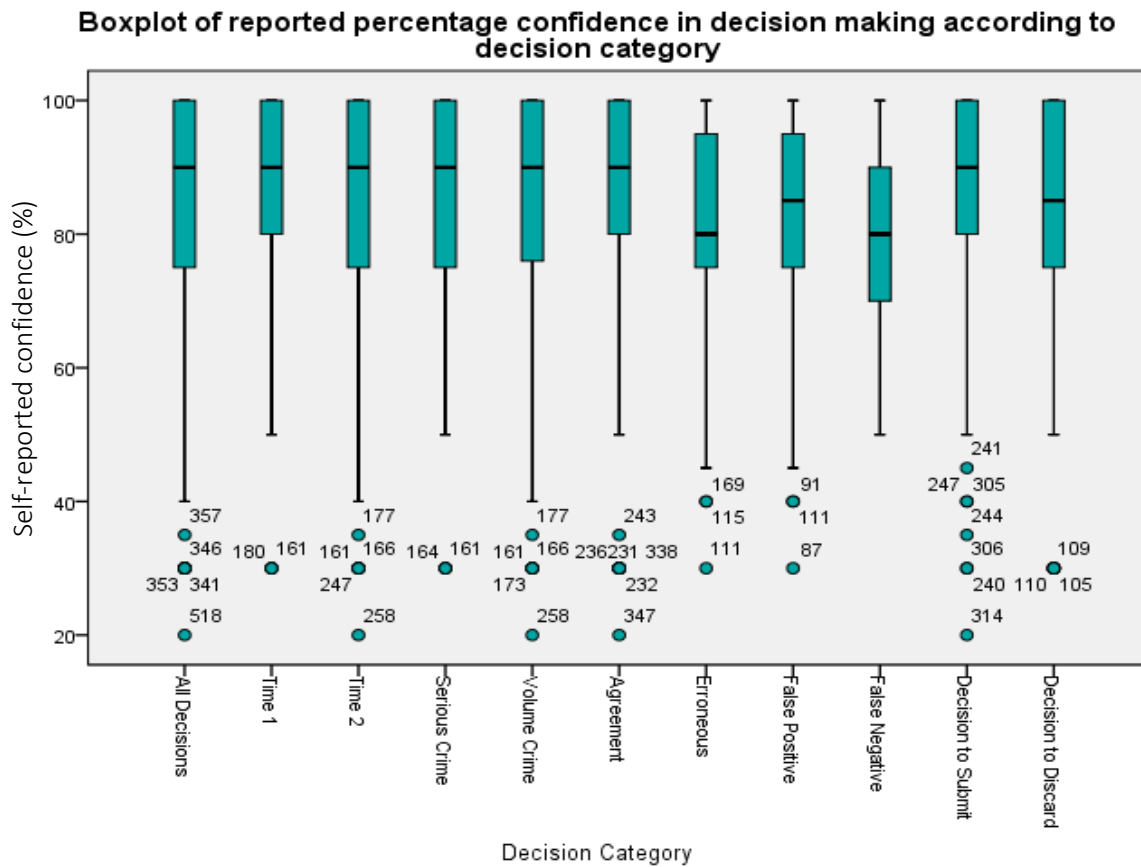
Descriptive statistics for confidence in all decisions made and according to each specified decision category are provided in Table 3.8.

Table 3.8 - Descriptive statistics in relation to self-reported confidence

		Descriptive Statistic						
		N	Range	Minimum	Maximum	Mean	Standard Deviation	Variance
Decision Category	All Decisions	520	80	20	100	84.32	16.43	269.96
	Time 1	260	70	30	100	85.73	14.90	222.01
	Time 2	260	80	20	100	82.92	17.75	314.99
	Serious Context	260	70	30	100	85.25	15.25	232.28
	Volume Context	260	80	20	100	83.39	17.52	306.95
	Submit	315	80	20	100	85.17	16.37	267.87
	Discard	205	70	30	100	83.02	16.48	271.70
	Outcome: Agreement	349	80	20	100	85.92	16.58	275.02
	Outcome: Erroneous	171	70	30	100	81.06	15.66	245.31
	Outcome: False Positive	113	70	30	100	82.10	16.64	277.00
	Outcome: False Negative	58	50	50	100	79.05	13.46	181.10

The spread of self-reported confidence levels for each decision category are also compared in the box plot included in Figure 3.12. Lower confidence can be seen to have been reported when an erroneous decision was made and also when this was a false negative type of error, as opposed to a false positive error.

Figure 3.12 - A comparison of self-reported confidence according to decision category



Categorised self-reported confidence data was analysed in order to identify any statistical differences in confidence according to the type of decision made. Data in relation to each decision category was found to be not normally distributed, violating the assumptions of a parametric test so Wilcoxon Signed Ranks tests were carried out in each case. The results of these statistical tests are summarised in Table 3.9.

Table 3.9 - Tabulated results of Wilcoxon signed ranks tests comparing self-reported confidence

Wilcoxon Signed Rank Test Results			
Comparison between:	N	P Value	Significant difference ( $p \leq 0.05$ )?
Reported confidence at Time 1 and Time 2	260	0.002	Yes, significantly more confident when making a decision at time 1 than at time 2
Reported confidence in serious and volume crime scenario	260	0.058	No significant difference
Reported confidence when submitting and discarding a fingerprint	205	0.05	Yes, significantly more confident when submitting a fingerprint than when discarding a fingerprint
Reported confidence when a decision was in agreement and erroneous in relation to the ground truth	171	0.00	Yes, significantly more confident when making a decision that was in agreement with ground truth
Reported confidence when a decision was false positive and false negative	58	0.00	Yes, significantly more confident when making a false positive decision than making false negative decision

Self-reported confidence according to experimental fingerprint

All confidence values were divided according to experimental fingerprint to determine the spread of confidence across the fingerprint set.

A box plot of self-reported confidence ratings according to the experimental fingerprint is provided in Figure 3.13.



Figure 3.13 - Boxplot illustrating self-reported confidence according to experimental fingermark

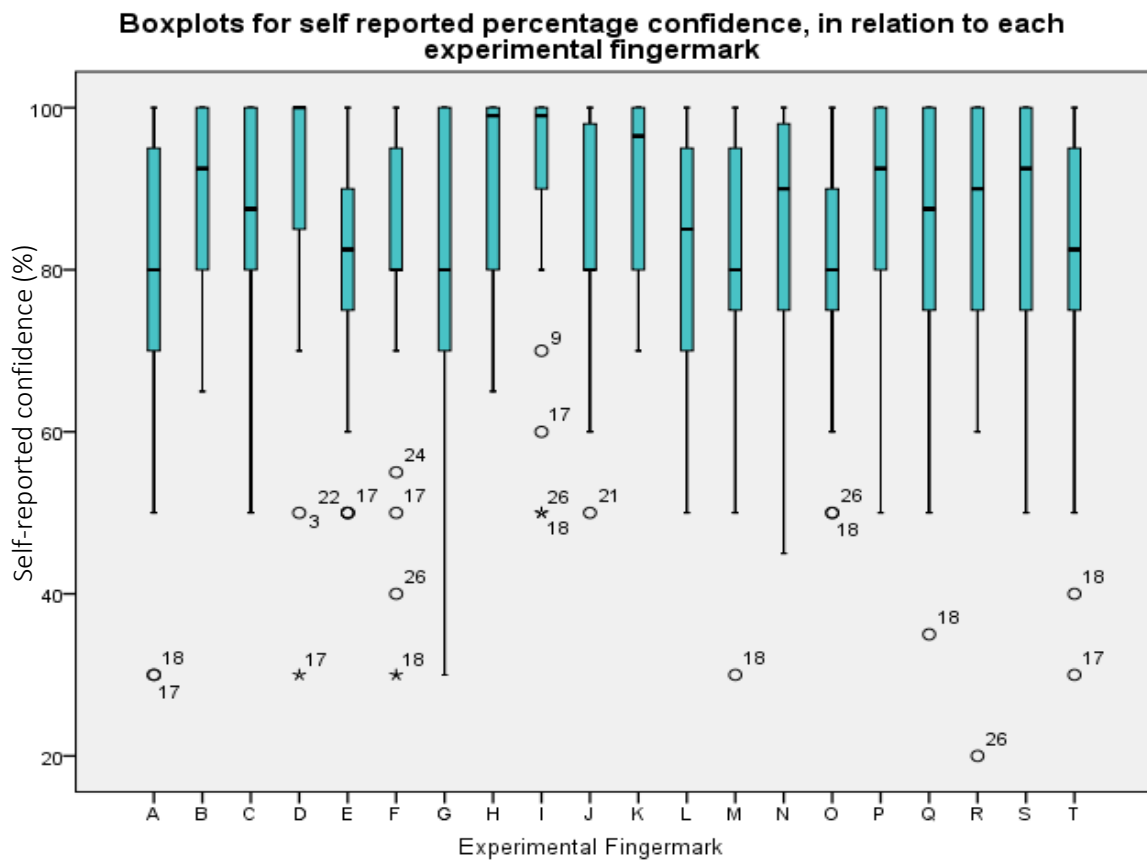


Figure 3.13 shows the mean confidence rating given in relation to each fingermark, showing a small variation in average confidence according to fingermark, with mean values falling within a 20% range.

Self-reported confidence according to practitioner

Confidence values were divided according to practitioner to examine the data for individual differences in self-reported confidence.

A boxplot of self-reported confidence rating according to participating practitioner is provided in Figure 3.14.

Figure 3.14 - Boxplot illustrating self-reported confidence according to practitioner

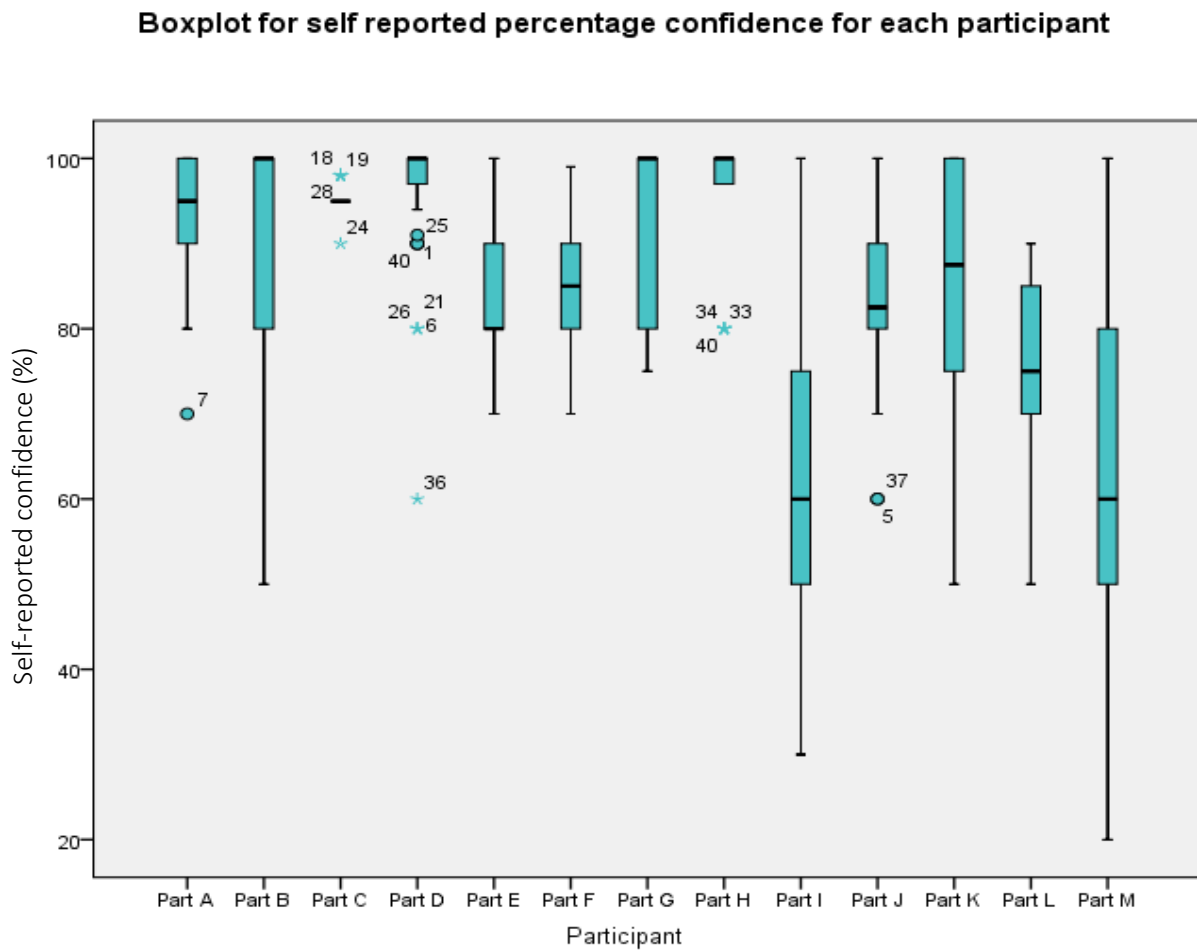


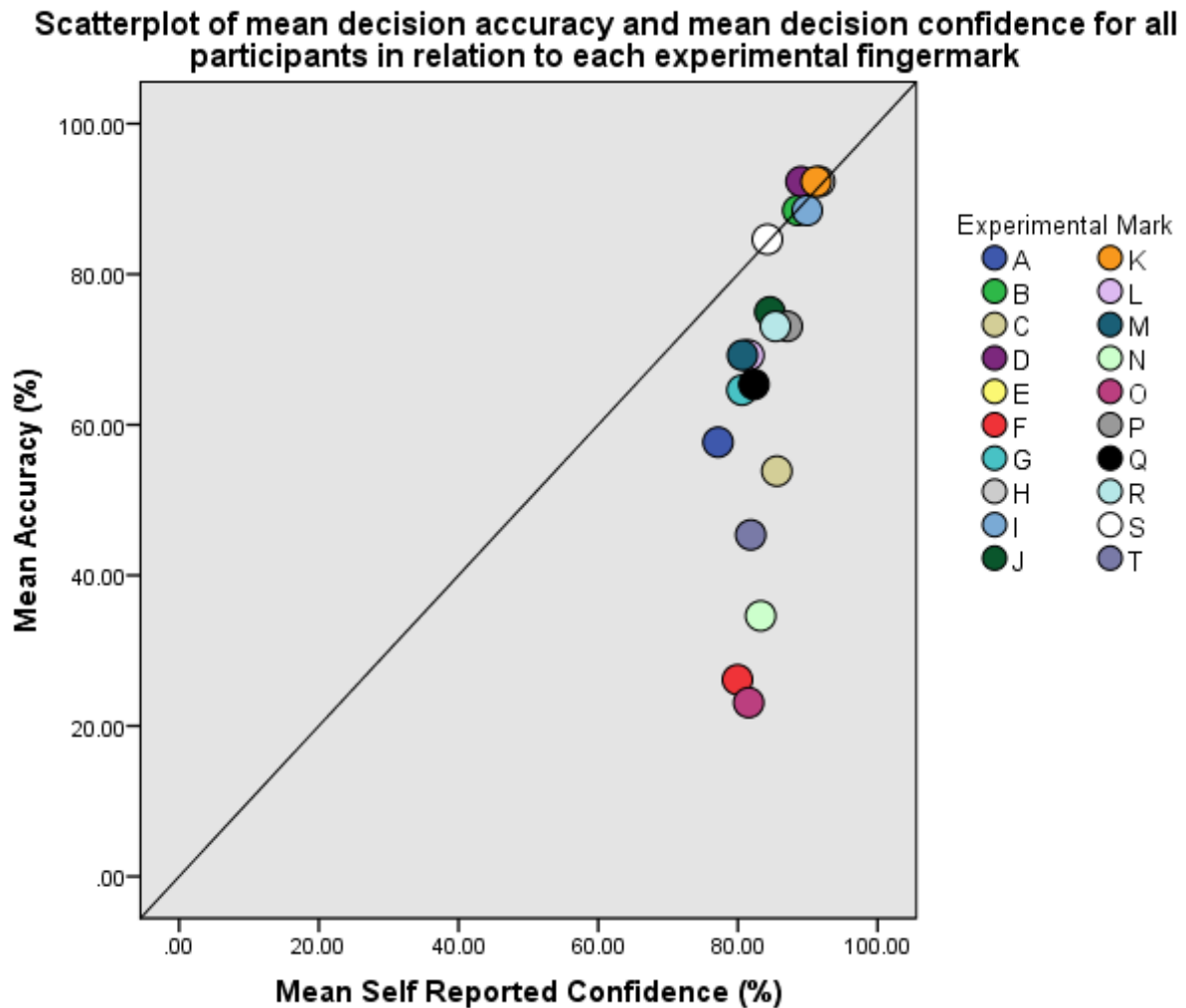
Figure 3.14 illustrates the individual differences present in relation to self-reported confidence. There is a much higher degree of variation in confidence determinations made by each practitioner than according to the fingerprint being considered.

Assessment of the relationship between confidence and accuracy (calibration)

Categorical accuracy results of 'agreement' or 'erroneous' decisions were converted to numerical data by scoring 'agreement' decisions as 100% correct and 'erroneous' decisions as 0% correct. Mean accuracy was calculated for all decisions made in relation to each fingerprint and by each participant.

Figure 3.15 shows mean accuracy plotted against mean confidence for all decisions made in relation to each fingerprint.

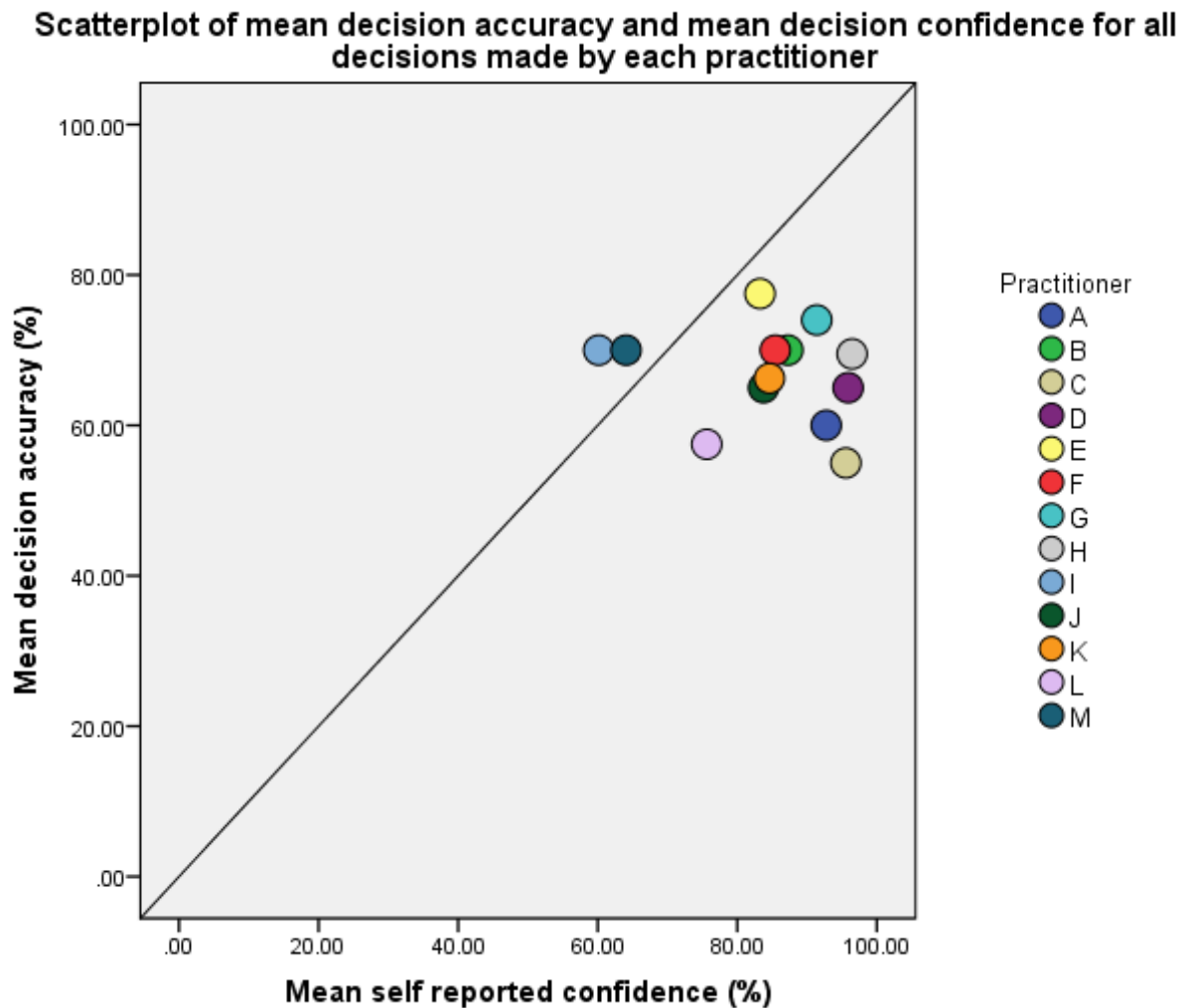
Figure 3.15 - Scatterplot of mean accuracy against mean decision confidence for decisions made in relation to each experimental fingermark



Within Figure 3.15 the line plotted at X=Y represents calibrated mean confidence, where a stated mean confidence level is in agreement with decision-making accuracy. Points below this line demonstrate overconfidence and points above it show under-confidence in decision making. As such a very slight positive visual correlation can be observed between confidence and accuracy in decision making. The marks upon which accuracy was, on average, the highest (marks B, D, I, K, and P) were also those about which practitioners, on average, reported the highest confidence in sufficiency decision making. In relation to these marks practitioners exhibited ideal calibration as they were neither over nor under-confident. Practitioners were overconfident in their decision making in relation to the majority of fingermarks and did not display under-confidence in decision making when results were considered according to experimental mark.

Figure 3.16 shows mean accuracy plotted against mean confidence for all decisions made by each of the practitioners.

Figure 3.16 - Scatterplot of mean accuracy against mean decision confidence according to practitioner



Again, in Figure 3.16 the line plotted at X=Y represents calibrated mean confidence, where a stated mean confidence level is in agreement with decision-making accuracy. Points below this line demonstrate overconfidence and points above it show under-confidence in decision making. Here there is no observable correlation between confidence and accuracy for each practitioner. Two practitioners (I and M) were, on average, under-confident in their decision making, while all other practitioners exhibited overconfidence.

Summary of objective 3.3. results

Practitioners were, overall, found to be overconfident in their fingerprint submission decision making as hypothesised. There were differences in reported confidence levels across practitioners according to the occasion upon which marks were presented, and whether the decision made was to submit or discard the mark. Practitioners were more confident when making a correct decision in terms of the

usability determination of an examiner, and were more confident when submitting a poor quality mark than when discarding a good quality mark.

#### 3.4.5. Objective 3.4: An assessment of inter-laboratory consistency in fingerprint submission decision-making

Results were achieved through the comparison of data presented within Objectives 3.1-3.3 in relation to Laboratory A with data gleaned in relation to the Metropolitan Police Service Serious Crime Fingerprint Recovery Laboratory, which is presented, in part in Earwaker et al. (2015). This comparison was carried out in order to establish the extent to which there was similarities in the efficiency of the fingerprint submission process between the two metropolitan laboratories. Data were compared in order to meet the sub-objectives outlined in section 3.2 (overall efficiency, the effect of crime contextual information, and the relationship between confidence and accuracy in decision making).

Objective 3.4.1. Comparison of the overall efficiency of fingerprint sufficiency decision-making between Laboratory A and the Metropolitan Police Service Fingerprint Enhancement Laboratories

Comparison of overall decision outcomes

Overall a total of 520 decisions were made by Laboratory A practitioners (2 decisions made in relation to each of the 20 experimental fingerprints by each of the 13 participants) and 440 decisions were made by Metropolitan Police Service practitioners (2 decisions made in relation to each of the 20 experimental fingerprints by each of the 11 participants). All results have been calculated in terms of percentages and rates so to enable a direct comparison between the two Police Forces, whilst allowing for the differences in participating practitioner numbers between the two laboratories.

66% of submission decisions made by both of the laboratories were in agreement with the 'ground truth' examiner assessment of the usability of the fingerprints. The proportion of the remaining 33% of erroneous decision outcomes is different for each laboratory. Laboratory A has a larger proportion of false positive decision outcomes (22% of decision outcomes were poor quality fingerprints submitted), than false negative decision outcomes (12% of decision outcomes were good quality fingerprints discarded), whereas the MPS lab has an equal proportion of false positive and false negative erroneous decisions (12% each). A Chi square test showed that the relationship between police force and decision outcome is statistically significant ( $\chi^2, 2 = 7.202, p < 0.05$ ). The Cramers V test statistic was 0.087 which indicates that the relationship between the two variables, however, was weak. The category that was most underrepresented was false negative decision outcomes within

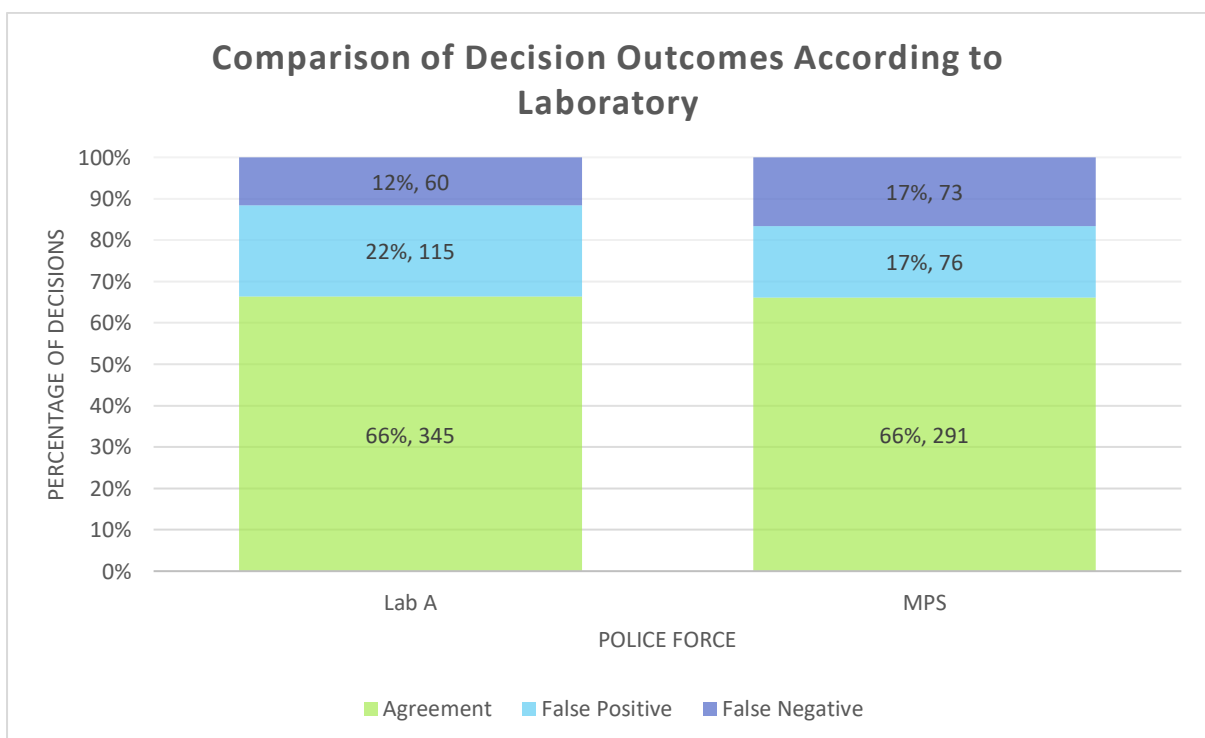
Laboratory A and the category that was most overrepresented was false positive decision outcomes within the MPS laboratory. The contingency table in relation to this analysis is provided in Table 3.10.

Table 3.10 - Contingency table for Chi Square test of relationship between laboratory and decision outcome

			Laboratory		Total
			Lab A	MPS	
Outcome	Agreement	Count	345	291	636
		Expected Count	344.5	291.5	636
	False Neg	Count	60	73	133
		Expected Count	72	61	133
	False Pos	Count	115	76	191
		Expected Count	103.5	87.5	191
Total		Count	520	440	960
		Expected Count	520	440	960

Figure 3.17. provides a graphical comparison of the overall decision outcomes generated by each of the laboratories during the course of the study.

Figure 3.17 - Comparison of overall decision outcomes according to laboratory



### Comparison of descriptive statistics

The descriptive statistics presented in Table 3.11. were calculated for the Metropolitan Police Service fingerprint laboratory data so as to provide points of comparison against the Laboratory A Police data points previously stated in the present chapter.

Table 3.11 - Comparison of descriptive statistics in relation to all decisions made by MPS and Laboratory A practitioners

	Lab A	MPS
False positive error rate	43.46%	33.19%
False negative error rate	22.31%	33.18%
False positive discovery rate	35.87%	34.08%
False negative discovery rate	28.29%	33.64%
Sensitivity	77.79%	66.82%
Specificity	56.54%	65.45%

The descriptive statistics calculated demonstrate that Laboratory A had a higher sensitivity in relation to the experimental marks, but a lower specificity than MPS. This means that laboratory A submitted a higher proportion of the identifiable marks, but discarded a lower proportion of insufficient marks, overall. Laboratory A also had a higher false positive error rate and false positive discovery rate than MPS, but a lower false negative error rate and false negative discovery rate, meaning that they submitted a higher proportion of insufficient marks, but discarded a lower proportion of identifiable marks. Overall, these values suggest a lower submission threshold within Laboratory A than the MPS laboratory.

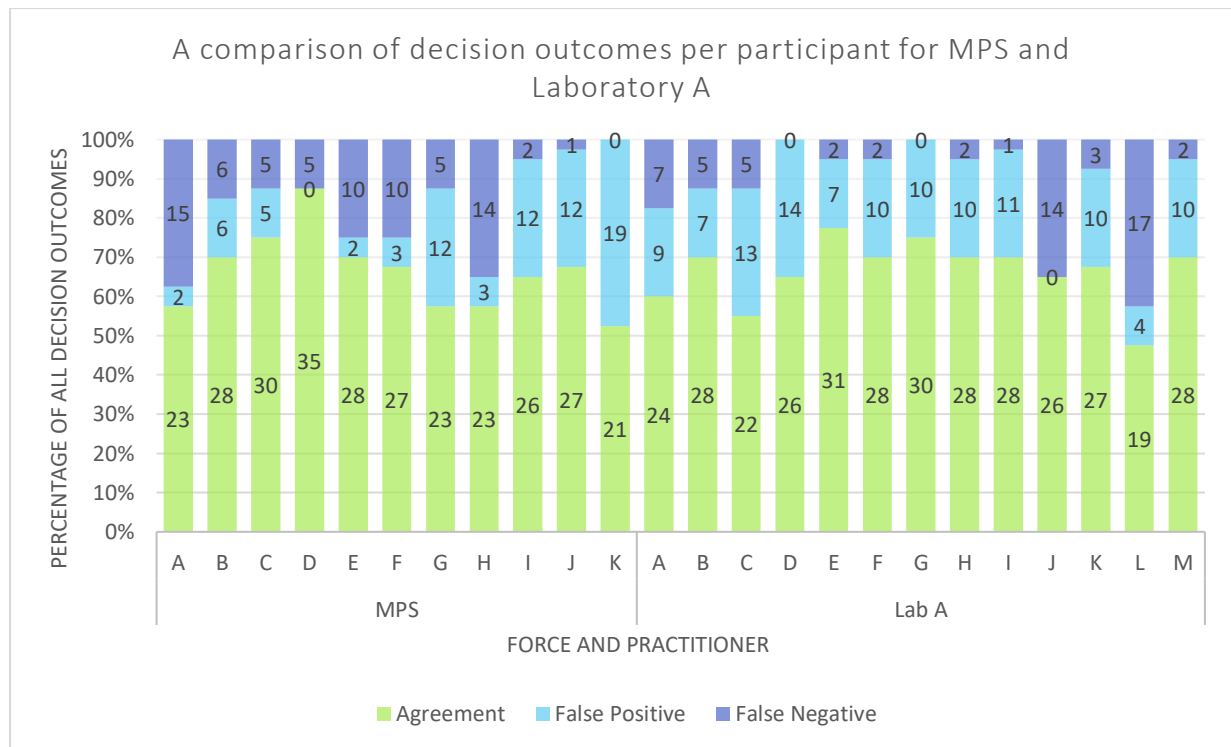
### Comparison of decision outcomes per fingermark

Chi squared tests were carried out to assess for statistically significant decision outcomes between the two laboratories in relation to each fingermark. The decision outcomes for all but one fingermark were found not to be significantly different between the two laboratories. The result of a Chi squared test (N=48, p=0.001) showed a statistically significant difference in decision outcomes in relation to fingermark Q, with a significantly higher proportion of erroneous (false negative) decisions made in relation to this fingermark within the MPS laboratory.

### Comparison of practitioner variability

The profile of practitioner decision outcomes for both laboratories are combined for comparison in Figure 3.18.

Figure 3.18 - Comparison of practitioner decision outcome profiles of MPS and Lab A



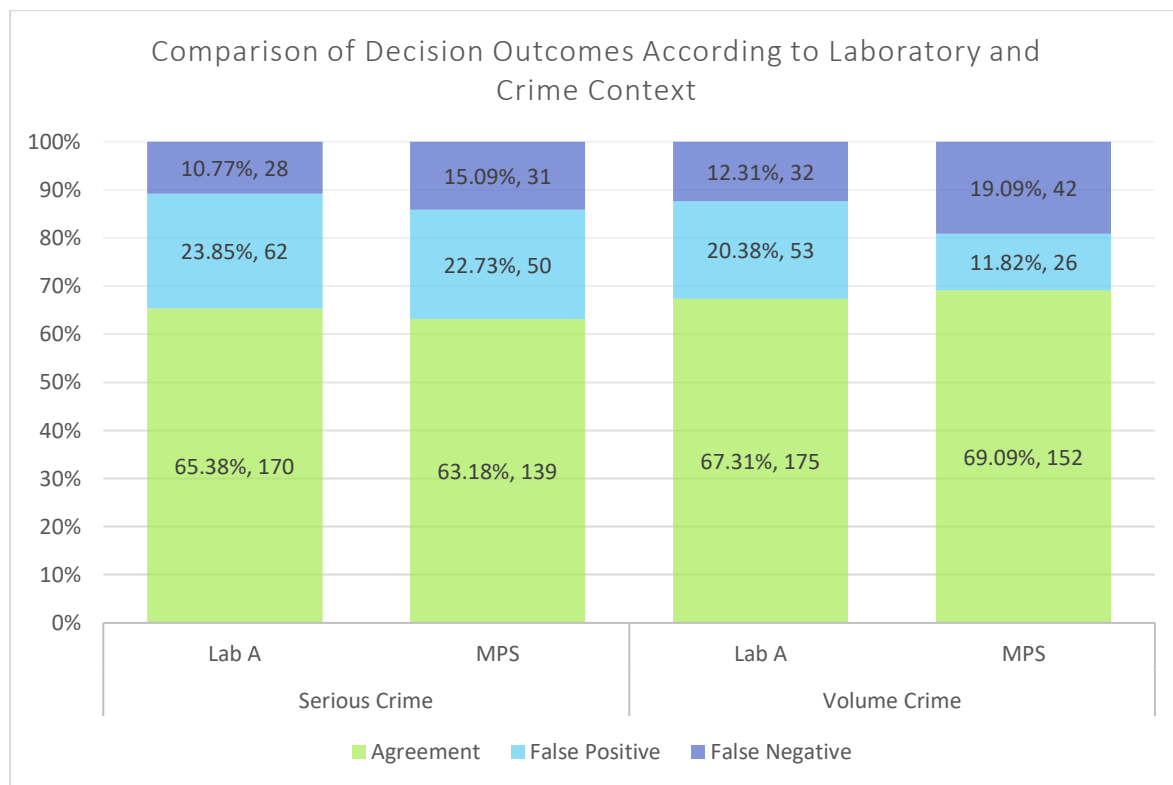
Whilst practitioners from MPS can be seen (in Figure 3.18) to have a slightly higher proportion of false negative decision outcomes, which is reflected in the overall comparison of decision outcomes, the profile of variation amongst practitioners is similar between the two forces. This highlights that errors are occurring across the board and that it is not certain practitioners that are responsible for erroneous decisions. The similarities in the practitioners of the two forces suggest that this pattern may be one that can be extrapolated to other groups of practitioners in other laboratories.

Objective 3.4.2: comparison of the effect of contextual information on fingerprint sufficiency decision making

Figure 3.19 illustrates the difference in the proportion of decision outcomes according to the crime context provided, for the two laboratories.



Figure 3.19 - Comparison of decision outcomes according to laboratory and crime context



The effect of the crime category on decision outcomes can be seen to differ between the two police forces. In the case of Laboratory A there was found to be no significant difference between the decision outcomes given when the experimental fingerprints were presented in a serious or a volume in context (a Chi – Square test showed that the relationship between crime context and decision outcome is not statistically significant ( $\chi^2, 2) = 1.043, p= 0.593$ ). Whereas in the case of MPS there was found to be a statistically significant difference between decision outcomes according to the context presented. The results of a Chi Squared test in this case, showed that there was a significant difference between the expected and observed outcomes when the decision outcomes were divided according to the crime category ( $\chi^2, 2) = 9.817, p < 0.01$ ). The Cramer V test statistic was 0.149 which indicates that the relationship between the two variables was, however, weak (Earwaker et al. 2015). As illustrated in Figure 3.19 within the MPS serious crime category false-negative decisions were underrepresented and false-positive decisions were overrepresented, whereas in the volume crime category the reverse was found.

Data was analysed to establish whether there was a statistically significant difference in the make-up of decision outcomes achieved by each police force for each of the two crime contexts. When fingerprints were presented in the context of a serious crime a Chi Square test showed there to be no significant difference in the make-up of decision outcomes between the two forces ( $\chi^2, 2) = 1.223, p= 0.542$ ). However, when the fingerprints were presented in a volume context the decision outcomes

of the two forces were significantly different ( $\chi^2,2=7.711$ ,  $p<0.05$ ). The Cramer V test statistic was 0.127 which indicates that the relationship between the two variables was, however, weak. Within the MPS false negative decision outcomes were overrepresented within the data whilst false positive decision outcomes were underrepresented, when the fingermarks were presented in the context of a volume case.

#### Comparison of decision consistency

Further interrogation of decision outcome data was carried out in order to establish whether there were differences in the consistency of decision making between the two laboratories. A comparison of the consistency of decision making of the two laboratories is tabulated in Table 3.12.

Table 3.12 - Tabulated comparison of decision consistency between MPS and Lab A

Aspect of decision consistency	Laboratory	
	MPS	Lab A
Total percentage of decision pairs that were inconsistent	25%	20%
Percentage of all decision pairs that were inconsistent with decision to submit mark made in serious context and decision to discard mark made in volume context	20.45% (81.82% of inconsistent decision pairs)	12.69% (63.46% of inconsistent decision pairs)
Percentage of all decision pairs that were inconsistent with decision to submit mark made in volume context and decision to discard mark made in serious context	4.55% (18.18% of inconsistent decision pairs)	7.31% (36.54% of inconsistent decision pairs)

There was not a statistically significant difference between the proportion of consistent and inconsistent decision pairs between the two laboratories (a Chi – Square test showed that the relationship between laboratory and consistent or inconsistent decision pair is not statistically significant ( $\chi^2, 2) = 1.720$ ,  $p= 0.190$ ). There was, however, a statistically significant difference between the proportion of inconsistent decision pairs in which the decisions were to submit a fingermark when it was presented in a serious context and to discard the same fingermark when presented in a volume context, and in which the decisions were to submit a fingermark when it was presented in a volume context and to discard the same fingermark when presented in a serious context. ( $\chi^2,2=4.559$ ,  $p<0.05$ ). The Cramer V test statistic was 0.206 which indicates that the relationship between the two

variables was, however, weak. The contingency table in relation to this analysis is presented in Table 3.13.

Table 3.13 - Contingency table for Chi Square test for difference in type of decision inconsistency according to laboratory

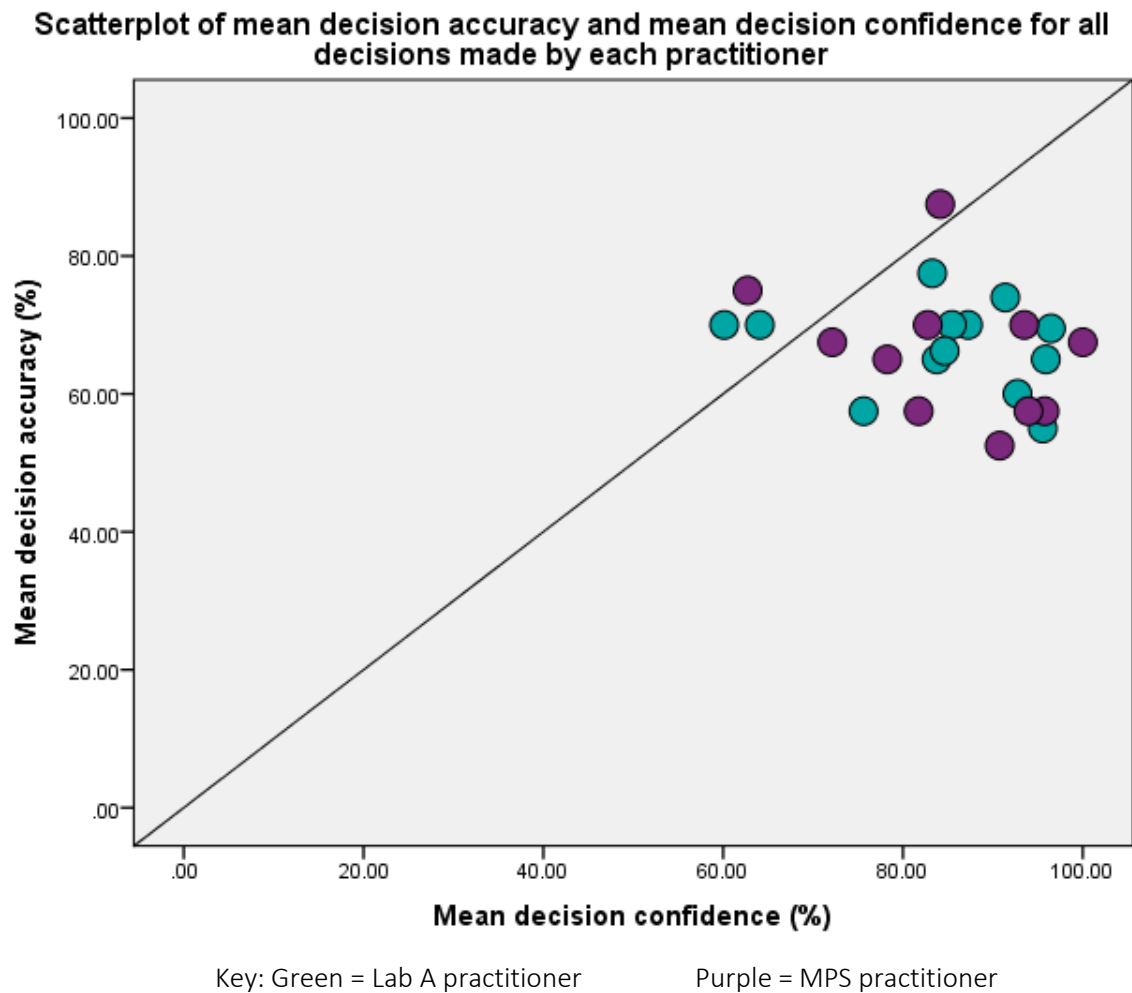
			Laboratory		Total
			Lab A	MPS	
Type of inconsistency	Submit in serious & discard in volume	Count	33	45	78
		Expected Count	37.9	40.1	78
	Submit in volume & discard in serious	Count	19	10	29
		Expected Count	14.1	14.9	29
Total		Count	52	55	107
		Expected Count	52	55	107

Within the MPS laboratory there were a higher proportion of decision pairs changing to reflect submitting in a serious but discarding in a volume case relative to Laboratory A, whilst the opposite was found to occur in relation to discarding marks in the serious context and submitting them in the volume context.

Objective 3.4.3: Comparison of the confidence and calibration of practitioners

The overall calibration (mean percentage accuracy plotted against mean self-reported confidence rating) of each participating practitioner from both laboratories is plotted in Figure 3.20.

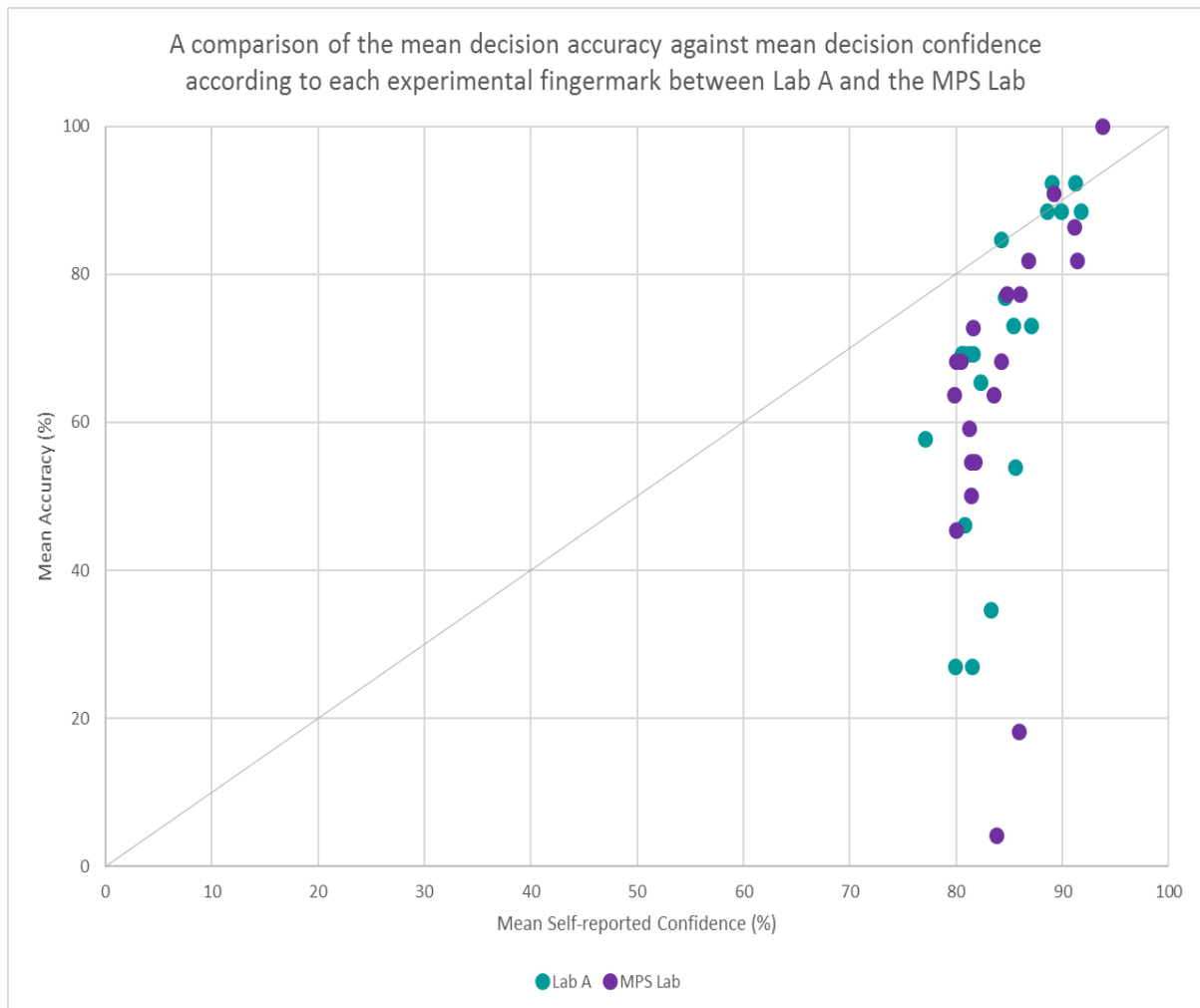
Figure 3.20- Mean decision accuracy and mean decision confidence for each practitioner from both laboratories



The majority of practitioners from both laboratories demonstrated overconfidence in their decision making during the experimental task (the data points in relation to these practitioners fall below the line of perfect calibration). Two practitioners from each laboratory displayed under confidence in their decision making.

Figure 3.21 shows mean percentage accuracy plotted against mean confidence for decisions made by both laboratories in relation to each experimental fingerprint.

Figure 3.21 - Comparison of collaboration per experimental fingermark



For both laboratories a slight directionality can be seen to be occurring in the relationship between confidence and accuracy. The fingermarks upon which the practitioners are on average most accurate in their sufficiency decision making are those upon which they report, on average, to be most confident. There is a very slight overall trend for practitioners to be more confident in their decision making the more accurate they are according to the fingermark in question, although this is a weak visual relationship. Overall the range of confidence and accuracy can be seen to be similar for both forces.

#### Summary of Objective 3.4. results

Objective 3.4. set out to compare the results of the analysis of fingermark submission decision making within the Metropolitan Police Service (Earwaker et al. 2015) with those presented in relation to another UK metropolitan laboratory (Laboratory A), in order to establish the extent to which the findings of Earwaker et al. (2015) were replicated outside of the MPS within a metropolitan police force that procures training from the College of Policing. The comparison established a similar level of

erroneous submission decisions between the two laboratories, but a difference in the effect of crime type on decision making performance, with the MPS laboratory appearing to adopt a higher submission threshold in cases of volume crime than Laboratory A.

### 3.5. Summary of key findings

The key findings of this study to date are summarised as follows:

Objective 3.1: Efficiency of fingerprint sufficiency decision-making within Laboratory A

- 33.65% of the practitioner submission decisions made during this study were erroneous (in relation to examiner usability determinations about ambiguous fingerprints). 34.29% of these were false negative decisions; 60 identifiable fingerprints were discarded during this study.
- Practitioners made more errors in relation to fingerprints of insufficient quality than in relation to sufficient quality marks.
- There was a variation in decision outcomes according to experimental fingerprint.
- Erroneous decisions were made by all practitioners. Whilst there was some variation in decision outcomes, the errors made were divided between practitioners and were not due to a small minority of practitioners.

Objective 3.2: Effect of contextual information on sufficiency decision-making within Laboratory A

- There was no statistically significant difference in decision outcomes according to the case context in which the fingerprints were presented. However, there were inconsistencies in decision making in relation to the same marks when the crime context was changed; a higher number of inconsistent decisions were to submit a mark when it was presented in a serious context, and to discard the same mark when it was presented in a volume context, as opposed to the other way around.
- There was variation in the decision consistency and type of decision inconsistency observed according to the experimental fingerprint.
- All practitioners were mostly consistent and correct in their decision making although there were some individual differences in decision inconsistency.

Objective 3.3 – The relationship between self-reported confidence and decision accuracy within Laboratory A

- Practitioners reported themselves to be less confident when discarding fingerprints, making an erroneous decision, and making a false negative decision.
- The majority of practitioners were overconfident in their sufficiency decision making.

- Practitioners were overconfident in their submission decisions made in relation to all but six of the experimental fingermarks. Practitioners were more confident in relation to these six marks, and these were the six with the highest decision accuracy.

Objective 3.4 – Comparison between mark submission within Laboratory A and the Metropolitan Police Serious Crime Laboratory

Objective 3.4.1: Comparison of the efficiency of fingermark sufficiency decision-making

- The percentage of 'correct' decisions (according to examiner usability determination) was the same for both forces. The two laboratories therefore also made the same proportion of erroneous decisions (33%).
- The erroneous decision of the two forces have a different make up. Laboratory A made a higher percentage of false positive decisions (22%) and a lower percentage of false negative decisions (12%) than the MPS laboratory, which showed an even divide of error types at 17% each.
- Decision making outcomes per fingermark showed similarities between the two forces.
- There are similar individual differences between practitioners in both forces.

Objective 3.4.2: Comparison of the effect of contextual information on sufficiency decision making

- There is a significant difference in the effect of crime context on decision outcomes between the two laboratories.
- In the case of the Laboratory A there was no significant difference in decision outcomes for the two crime contexts, but there was a significant difference found in decision outcomes according to context within the MPS laboratory.
- The MPS laboratory was less consistent in its decision making than Laboratory A (although this was not found to be a statistically significant difference).
- Where there was inconsistency in decision making the MPS laboratory submitted marks in a serious case and discarded the same mark in relation to a volume case more than the Laboratory A, and discarded marks in a serious case whilst submitting the same mark in a volume case less than the Laboratory A.

Objective 3.4.3: Comparison of the relationship between self-reported confidence and decision accuracy

- Similarities in mean calibration were noted for practitioners from both laboratories.

- The majority of practitioners from both laboratories were found to be overconfident in their decision making.
- There was no correlation found between confidence and accuracy according to practitioner for either laboratory.
- There were similarities in calibration according to fingermark demonstrated by the two laboratories.
- In the case of both laboratories there was a weak visual correlation between accuracy and confidence according to experimental fingermark.
- Both laboratories demonstrated good correlation in relation to the fingermarks upon which they been most accurate, and the identity of these marks was common to both laboratories.
- Both laboratories were over confident in their decision making in relation to the majority of fingermarks.

### 3.6. Discussion

#### 3.6.1. Objective 3.1: Assessment of the efficiency of fingermark submission decision making within Laboratory A

The headline finding of this study is that 33.65% of the fingermark submission decisions made by practitioners in relation to the experimental set of fingermarks provided were erroneous in relation to examiner determinations of mark usability. During the study, practitioners erroneously decided to discard 60 identifiable fingermarks. This is, perhaps, the most important finding of this research as it suggests the potential for the loss of fingerprint evidence in casework, which could have led to the identification of, or the acquisition of intelligence against, a person of interest. Indeed, 28 of these fingermarks were discarded when practitioners were asked to view the submission task as if the fingermarks were those developed upon evidence in relation to the serious case of a murder. Whilst this may seem to be a concerning finding it should be born in mind that the fingermarks included within the experimental set were selected as it was felt that they were examples of ambiguous, or 'borderline' marks which would provide a decision making challenge to the practitioners. The marks were not (and neither were they intended to be) representative of the cross section of fingermarks that would be developed during real life casework. Subsequently the error percentages calculated here should not be extrapolated to provide a casework error rate. Indeed, care should be taken with the reporting of error rates within forensic science which are based upon such empirical studies (Christensen et al, 2014). However, whilst these marks were intentionally ambiguous in order to provide a challenging exercise for practitioners, they cannot be said to be the most ambiguous or challenging marks which could be encountered during casework. A total of five fingerprint examiners



were employed to provide a usability determination in relation to each of mark considered for inclusion within the experimental set. The 20 marks ultimately selected to form the experimental set were those upon which agreement in usability had been reached by all of the five examiners. It is important to note that agreement by all examiners was only reached in the case of these 20 marks; the examiners did not all agree upon the sufficiency of the remaining 20 fingermarks put forward for use in this study. This in itself is an interesting finding in relation to the variation of examiner mark quality assessment. Inconsistency between the judgements of fingerprint examiners has been demonstrated experimentally (Dror et al., 2011). Langenburg et al. (2009) found the reproducibility (consistency between examiners) and repeatability (ability of an examiner to gain consistent results in relation to the same print) of the final decision of the ACE-V process to vary according to the quantity of information that was present within a fingermark, supporting the idea that variability was found to occur within this study when prints contained less information. The reproducibility of examiners in relation to the experimental fingermarks therefore suggests that the selected marks have a reasonable quantity of information within them and are not the most challenging, for quality determination purposes, that could possibly be encountered within casework. Consequently, the results of this study cannot be considered to only apply to outlying, or extreme marks and so can be said to be more ecologically valid. The marks are, however, ambiguous, requiring practitioners to make decisions under uncertainty.

Both false positive errors (submitting marks of too poor a quality) and false negative errors (discarding good quality marks) were made by the practitioners during this study. This would suggest that there is no clear cut threshold difference between the sufficiency determination of practitioners and examiners. Had all errors been of one type this would have been strong evidence to suggest that a considerably different submission threshold was being employed by the two groups. Results did, however, show that a higher proportion of the errors made during this study were false positive than false negative, meaning that it was more common for practitioners to err on the side of caution and suggesting a possible tendency towards a lower submission threshold for practitioners than the usability threshold of the examiners, although the similar quantity of false negative errors to false positive errors would suggest that practitioners and examiners may instead be using different criteria for determining suitability, or may be using a similar criteria but have differing abilities in its application. The lack of a clear cut threshold difference may suggest that practitioners are simply submitting and discarding the wrong marks.

Signal detection theory highlights two distinct aspects of decision making in ambiguous cases; the ability of a decision maker to detect a signal (in this case recognise the presence of quality indicators within a fingermark) and the threshold at which that individual determines this signal to be significant

(the quantity of quality indicators or the level of quantity required for a submission decision to be made) (Phillips et al., 2001). This poses an interesting question in relation to the experimental findings; are practitioners submitting the wrong marks to examiners because of differences in their ability to detect signals, or features, within the fingerprint or are they adopting the incorrect threshold in relation to that of the examiner? As there is not an overall threshold difference it would seem that it may be the case that either the practitioners have a different ability to detect the signals present within the marks or they place different values on some signals to examiners resulting in apparently different differences in judgement. Further analysis of variation between experimental fingermark and between practitioners may assist in further generating such a hypothesis.

There are a number of theories that may explain a difference in either the detection of features (signals) or decision threshold between the fingerprint examiners and the laboratory practitioners, such as the finding that experts perform better at pattern recognition tasks (Chi, 2006) or that examiners work to a 'winner takes all' decision making threshold (Charlton, 2010) meaning the quantity of information required to make a decision about a fingerprint may vary between an examiner and practitioner due to the greater emphasis on training and feedback in the 'analysis' process provided to examiners. The finding that examiners are able to identify a higher number of minutiae within a fingerprint (Langenburg, 2004, Schiffer & Champod, 2007), may provide support for there being differences in signal detection between the two groups in the cases in which a false negative decision was reached, but this explanation alone does not account for the false positive submissions. It is also important to consider that differences in methodology may have an effect as fingerprint examiners carry out their 'analysis' with reference to a control print (as they did in the present study) resulting in a difference in judgement to a practitioner viewing the print in isolation (Fraser-Mackenzie et al., 2013), although such a difference is also true of standard workflow.

However, the results of the present study suggest there is not a straight-forward difference in perceived fingermark sufficiency threshold between the examiners and the laboratory practitioners. Practitioners are not just sending marks at too high a threshold (missing good marks) or sending marks at too low a threshold (sending poor quality marks); they are sending and discarding the wrong marks. This suggests, rather than a clear cut case of a threshold difference between the laboratory practitioners and fingerprint examiners, the wrong information is being used or treated in the wrong way in each decision. The lack of a clear threshold goes against the idea that there is a single factor that is causing a key difference such as the ability to identify minutiae, and suggests more that the decisions are being made in different ways by the two groups, highlighting a key area for future research to ascertain the differences in laboratory practitioner and fingerprint examiner decision making, which will be further explored in Chapter 4. No clear fingermark submission policy was stated

by the laboratory manager and it may be the case that the generation of such a policy in line with the requirements of the fingerprint examiner would be successful in increasing the accuracy of decision-making according to the needs of the examiners, although the subjective and experience-based nature of expert decision making (Ulery et al., 2013) highlights further challenges inherent in this task. Further investigation is needed to determine the cognitive ability of practitioners and examiners to uphold any policies made in order to be effective at providing a more efficient fingerprint recovery process.

### 3.6.2. Objective 3.2: The effect of contextual information on practitioner submission decision making within Laboratory A

There was no significant difference found between the overall make up of outcomes of the decisions made by the practitioners of Laboratory A when the context of the case in which the fingermarks were presented changed. This supports the findings of Schiffer and Champod (2007) who showed that context did not have an impact upon the number of minutiae counted by participants. This may suggest that minutiae count (or the detection of this quality signal (Phillips et al., 2001)) is of paramount importance in determining fingerprint sufficiency and that this task is, perhaps, unaffected by context.

However, upon further analysis of the data to determine the provenance of the decision outcomes through investigating decision consistency, it was found that, when the decision of a practitioner changed in relation to the same mark, it was more often the case that the directionality of this inconsistency was to submit the mark when it had been presented in a serious case, and to discard the same mark in a volume case, rather than to make the opposite decisions. This would suggest that the contextual information presented had some effect on decision making. Given that Schiffer and Champod (2007) found that contextual information had no effect on minutiae account it may be that this information is having a slight effect on submission threshold, rather than purely signal detection, i.e. practitioners are detecting the same number of minutiae irrespective of context, but are altering the number of minutiae that they consider to be the threshold for submission according to the differing utility values placed upon the evidence by the contextual information (Gittelsohn et al., 2013). The presence of a slight, but not clear cut, effect of contextual information is interesting, and further suggests the need for further research to ascertain the relative importance of signal detection and threshold in sufficiency decision making. These results suggest that there is value in further work into bias within fingerprint submission and wider forensic science, contrary to the recommendations of Champod (2014), as such research has the potential to be used to establish a greater understanding of the exact processes and vulnerabilities present within this application of decision making.

### 3.6.3. Objective 3.3: Assessment of the relationship between self-reported confidence and decision accuracy

The relationship between confidence and decision outcome

Statistical comparisons of self-reported confidence showed practitioners to be less confident when deciding to discard fingerprints than when submitting them, less confident when making an erroneous decision than a correct one, and less confident when the outcome of their erroneous decision was a false negative rather than a false positive. These findings may provide some evidence for the potential usefulness of confidence ratings as a decision tool as it appears that practitioners are being less confident when they are making erroneous decisions and when they are making, what could be considered to be, more costly errors (a false negative decision as opposed to a false positive decision). Indeed, confidence ratings have been discussed as an indicator of decision success within eye witness testimony (Wells et al., 2002). Eye witness line up identifications are a relevant parallel task to fingerprint sufficiency decision making as, whilst both occur within the forensic domain, they also both require the perceptual tasks associated with feature recognition and determining 'hit' thresholds. It may be the case that there would be benefit in the use of confidence ratings to provide a mechanism for flagging up challenging decisions upon which a second check by a different practitioner, or by an examiner would be beneficial.

However, it is important to put these findings into the context of the data pool resultant from the present study. Participants were also found to be more confident in their decision making upon the first presentation of the experimental fingerprints (T1) than during the second presentation (T2). This is, perhaps, an unexpected finding, as it has been found that practice (even without feedback) increases confidence in decision making (Paese & Snizek, 1991) which may have suggested that increased confidence would have been more likely to have been observed in relation to the second presentation of the fingerprints. It also suggests that there may be a factor other than accuracy influencing confidence in decision making, as no statistically significant difference in decision accuracy was found between the two presentation occasions. Equally it should be born in mind that the data pool in relation to false negative decision outcomes was considerably smaller (due to a lower number of false negative decisions being made during the study) and this may have skewed statistical analysis.

Calibration

When calibration was examined results showed a predominant theme of overconfidence in decision making. When analysed according to the mean confidence and accuracy of each participant all but two practitioners were found to be overconfident in their decision making and no relationship was seen to exist between confidence and accuracy of decision making. Further interrogation of the data sheds further light upon the outlying practitioners not found to be overconfident. Both had relatively

high levels of accuracy, but had the lowest mean self-reported confidence. It is interesting to note that both practitioners had the same, considerable amount of experience (14 years) of working in the laboratory, suggesting a positive relationship between expertise and calibration. This supports the finding that experience can improve calibration by reducing overconfidence (Murphy & Winkler, 1977, Garb, 1986). However, there was not a clear trend identified between calibration and experience, and there was one practitioner with more experience (18 years) who was found to be overconfident. Indeed, there is an argument that extensive experience of carrying out a task without being criticised may lead to overconfidence as it has been found that high self-belief and increased experience have been found to produce overconfidence in decision-making (Heath & Tversky, 1991, Armelius & Armelius, 1976).

When calibration was examined according to the mean decision accuracy and confidence ratings made by all practitioners in relation to each experimental fingerprint, a weak visual relationship between confidence and accuracy was found. Practitioners appeared to be slightly more confident in making sufficiency decisions in relation to fingerprints that they had been more accurate about. Indeed, practitioners demonstrated ideal calibration in relation to the six fingerprints about which their decision making had been most accurate (marks B, D, H, I, K, and S). Practitioners were overconfident in relation to all other fingerprints. This suggests, again, that the mark itself is important in sufficiency decision making. It would be interesting to further examine the fingerprints used in this study to gain a better understanding of the similarities and differences in the presence of quality features between marks about which high and low accuracy and calibration were found.

It would be interesting to further examine calibration from a decision theoretic perspective. Accuracy for the purposes of analysis to date was defined as 100% if the decision made was in agreement with the ground truth of the examiners, and 0% if the decision was erroneous. The mean of these accuracy values was used to establish calibration. It would be interesting, however, to alter the weighting placed upon the different decision outcomes possible within the study and to assess the effect that these varying utility values have on the calibration of practitioners. For example, from a decision theoretic perspective (Gittelsohn et al., 2013), it could be said that the discarding a good quality fingerprint is more costly than submitting a poor quality fingerprint. As such it may be that a higher accuracy rating could be assigned to false positive errors than false negative errors, and calibration could be calculated based upon this weighted accuracy level. It also may be the case that the utility values placed upon certain decision outcomes may vary according to the context of the case type, or due to the piece of evidence in question. For example, it may be considered more costly to make a false negative error in relation to the sole fingerprint on a murder weapon than one of many fingerprints left by an offender in a burglary. This is an interesting area for future consideration.

Finding overconfidence to be a predominant effect within fingerprint sufficiency decision making is not surprising. Charlton (2006) found that 'it is a sign of weakness in the fingerprint profession to display anything other than absolute certainty'. This may suggest that the culture of forensic science is partly responsible for overconfidence and that practitioners did not want to appear unconfident in decision-making tasks that are fundamental to their job role.

Being accountable for a judgement has been found to reduce overconfidence, in an attempt to avoid embarrassment if later found wrong (Tetlock & Kim, 1987). This effect may not have been observed here because of the strong cultural emphasis within forensic science to be highly confident and, perhaps, due to the study being carried out anonymously meaning that the practitioners knew they were not individually accountable for erroneous decisions.

Overconfidence can be problematic as it does not flag areas in which additional training or procedure would be beneficial, instead it suggests that laboratory staff feel comfortable in their decision-making ability, when this may not be the case, hindering improvement in an area of the role where proficiency is often not appropriately tested. Equally it could be problematic in court if a practitioner claims to be highly confident in their decision made to discard a fingerprint but a defence fingerprint examiner has been able to identify that mark, damaging the credibility of the practitioner as a witness and the credibility of other fingerprint evidence in the case.

#### 3.6.4. Objective 3.4: A comparison of fingerprint laboratory fingerprint submission decision making between the Metropolitan Police Service and Laboratory A

##### Objective 3.4.1: comparison of fingerprint sufficiency decision making

Comparison of the success of practitioner submission decision making in relation to the fingerprint examiner determined usability of challenging fingerprints demonstrated an overall similarity between the two UK laboratories. Both laboratories were accurate in their mark submission (in relation to the usability of the examiners) in 66% of decisions made. Such a similarity between the two laboratories featured in this study is interesting, as one force procures training for laboratory practitioners from the College of Policing, whilst the other provides purely in house training. The replication of the overall success of submission decision as found within the MPS laboratory (Earwaker et al., 2015) may suggest the efficiency of decision making in relation to borderline fingerprints as used in this study within other UK laboratories who also procure their training from the College of Policing. However, it may be that particular differences in practitioner training do not affect the outcome of fingerprint submission, rather that the broad similarities in the national and local training programmes can explain similarities in overall decision making. Equally it may be that there are other similar factors between the two

forces such as working relationships between the laboratory and bureau that have led to this similar outcome, in addition to both fingerprint departments carrying out the same overall role.

Similarities in decision success between the two laboratories were also shown in relation to individual fingermarks. This suggests that practitioners find that the task of mark submission is more challenging in relation to some fingermarks, and a key area of future research to better understand the naturalistic mechanism of mark submission being employed would be to seek to establish the components of these marks which make them challenging in relation to practitioner submission. This could help to identify problematic aspects of mark quality assessment upon which future training and cognitive feedback could focus (Shepherd & Zacharakis, 2002), and to isolate the key features of fingermarks upon which submission decisions are based. It may be that the application of Signal Detection theory would be beneficial to examine the use of certain signals and thresholds within mark submission (Phillips et al., 2001).

Comparison of decision outcomes on an individual practitioner level between the two laboratories demonstrated errors (in relation to examiner usability determinations) by all practitioners in both laboratories. This is a key finding as it highlights the ubiquitous nature of 'action errors' (unintended deviations from a goal) (Frese & Keith, 2015), or 'practitioner errors' (Christensen et al., 2014), during the subjective and interpretative process of fingerprint quality assessment due to the 'error prone heuristic processes of human cognitive apparatus' (Reason, 1990). Given the ubiquitous nature of error determined here, and commonly found difficulties in fingerprint quality assessment task amongst those provided with training (Fieldhouse & Gwinnett, 2016) it would seem that the existence of error in such tasks needs to be recognised, and openly discussed through a framework of error management (Frese, 1991).

Objective 3.4.2: comparison of the effect of contextual information on fingerprint sufficiency decision-making

MPS and Laboratory A demonstrated a statistically similar make up of decision outcomes when fingermarks were presented in a serious crime context, but the two laboratories differed significantly in their decision outcomes when the marks were presented in relation to volume crime ( $\chi^2, 2=9.817$ ,  $p<0.01$ ). Practitioners within the MPS laboratory appear to have reduced their submission threshold in the case of volume fingermarks whereas the submission threshold adopted by Laboratory A has remained consistent. This difference may be explainable by the distinct difference in routine casework profile between the two laboratories. The MPS laboratory featured within Earwaker et al. (2015) is one of a number of laboratories within the Metropolitan Police Service, and deals primarily with casework relating to serious category crime. Laboratory A, in contrast, is responsible for all fingerprint

visualisation work for the force (resulting in the routine undertaking of both serious and volume category crime casework). Given this key difference it may be the case that the volume crime context had a significant effect on the decision making of the MPS practitioners as this unfamiliar task may have led to more uncertainty in submission decision making and either a conscious or subconscious raising of their submission threshold to avoid false positive submissions. A subconscious increase in submission threshold could be considered to occur due to the increased reliance of heuristics given the increased ambiguity of the lesser encountered crime type (Kahneman et al., 1982), and this could therefore be described as the presence of a 'cognitive bias' as widely discussed within the forensic science literature (Dror et al., 2005, 2006, 2011, Charlton et al., 2010). An increase in decision threshold may, on the other hand, have occurred as practitioners believed that they needed to change their threshold in order to meet a different definition of decision success relevant to a volume crime. It may be the case that Laboratory A practitioners dealing with both types of crime routinely were not affected by such subconscious cognitive biases and were aware that a successful submission decision within their organisation was the same irrespective of crime type. The finding that contextual information had a greater effect on the submission decision making of MPS practitioners is supported by consideration of instances of inconsistency in individual practitioner mark submission. Where there was inconsistency in decisions made by practitioners in relation to the same mark on two separate occasions practitioners within the MPS laboratory submitted the mark in a serious case and discarded the same mark in a volume case more often than practitioners in Laboratory A, and discarded marks in a serious case whilst submitting the same mark in a volume case less than the Laboratory A.

Objective 3.4.3: comparison of self-reported confidence and calibration in fingerprint sufficiency decision making

There was a similar distribution of overall calibration (mean accuracy in decision making plotted against mean self-reported confidence in decision making) for practitioners in both laboratories. Neither confidence ratings nor accuracy were significantly higher for either laboratory. The majority of practitioners from both laboratories were, however, overconfident in their decision making. These findings would suggest that whilst individual practitioner confidence in their accuracy is not reflected in their actual accuracy in the task provided, organisational culture does not seem to be having an effect on practitioner self-reported confidence in submission decision making. The further use of confidence ratings within signal detection theory and ROC analysis would be a potentially fruitful extension of this study in order to increase understanding of the role of thresholds and cues (Phillips et al., 2001). It is important, however, to first ascertain the robustness of the confidence data generated as a predictor of accuracy, as in other fields, confidence has also been found to be a poor indicator of performance in a task (Ames et al., 2010).



### 3.7. Conclusion

The results outlined and discussed within this chapter illustrate potential inefficiencies within the mark submission process when dealing with fingerprints of ambiguous quality. Additionally, it has been shown that these findings may be replicated across laboratories in the UK. Crime contextual information has been shown to have an effect upon mark submission decision making, but only in the case of the introduction of an unfamiliar mark submission crime category. The success of ambiguous mark submission has been shown to vary according to the mark in question, and has identified similarities in success according to fingerprint within the two laboratories. Erroneous decision outcomes have been shown to occur as a result of decisions made by all practitioners, highlighting the ubiquitous nature of such error. Practitioners have also been found to be more confident than they have been in their submission decision accuracy. Such findings demonstrate the importance of further research to empirically increase understanding of the mechanisms of the submission decision made by practitioners. Chapter 4 of this thesis seeks to initiate this process through a qualitative investigation of practitioner reasoning in relation to mark submission decision.

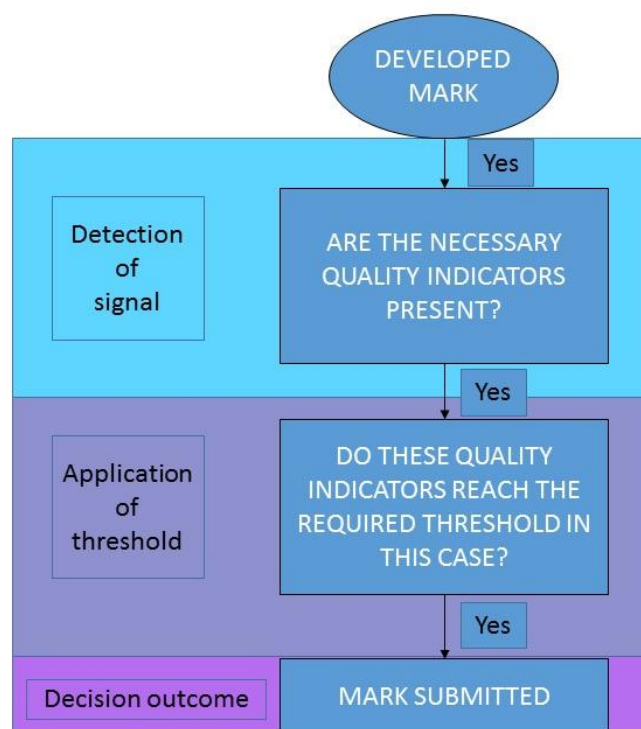
# Chapter 4 An investigation of self-reported fingerprint quality assessment decision rationale

## 4.1. Introduction

Initial research (Earwaker et al., (2015) and Chapter 3 of this thesis) demonstrates that there are inconsistencies between the fingerprint submission decision made by laboratory practitioners and the usability determinations made by fingerprint examiners, and that this is not an issue limited to one police force. A logical next step in the exploration of the practitioner submission decision is to consider how it is that submission decisions are being made by practitioners, how it is that usability decisions are being made by examiners, and to identify similarities and differences between these two decision mechanisms.

The concept of using the language of signal detection theory as a framework by which to describe the mark submission decision seems sensible as it allows discussion of the detection of quality indicators within a fingerprint separately from discussion of a threshold for submission employed (Phillips et al., 2011), as illustrated in Figure 4.1.

Figure 4.1 - A fingerprint practitioner decision to 'submit' described in the language of signal detection theory



By breaking down the decision into the two elements of signal detection theory it is possible to begin to consider the relationship between these two elements. Both need to be taken into consideration in order to fully understand the decision making process.

The research described in Chapter 3 of this thesis and Earwaker et al. (2015) provided some initial insight into the decision making process with regards to the decision thresholds of laboratory practitioners. This research found that the submission decisions made by laboratory practitioners do not appear to be being made at a different threshold to the usability determinations of fingerprint examiners (either generally submitting at a lower level and being more cautious or submitting at too high a threshold leading to missed marks). Instead it appears to be the case that the wrong fingermarks are being submitted and discarded by practitioners without any apparent correlation with a submission threshold. However, thresholds do appear to factor in to mark submission decision in some cases. Within empirical data gathered from participants employed within the Metropolitan Police Service Serious Crime Evidence Recovery Unit it was found that the threshold for the submission of borderline quality fingermarks to the fingerprint bureaux was lower in the case of a serious crime. This may have suggested either a conscious or subconscious reliance of case type to determine a submission threshold. However, when this relationship was considered in relation to the data gleaned from Laboratory A, some interesting similarities in threshold were apparent. The submission threshold remained constant for serious and volume crime within Laboratory A and this threshold mirrored the threshold of MPS when making decisions in relation to serious crime (the crime type most commonly encountered within this laboratory). This would suggest that in an unfamiliar circumstance where the degree of uncertainty was greater (dealing with volume crime) the practitioners increased their submission threshold. This may suggest that in uncertain situations crime type can alter submission thresholds.

Gaining a knowledge of practitioner's naturalistic thresholds and the factors that affect these when making a mark submission decision is important, as understanding and determining how to manipulate these thresholds is vital to ensure that mark submission occurs at an appropriate threshold according to the organisation, crime type, case, exhibit, and resources available (to name a few variables which may require alteration of the threshold). However, for the purposes of comparing practitioner submission decisions and examiner usability decisions it is the 'detecting quality signals' (to use the language of signal detection theory (Phillips et al., 2001)) aspect of decision making that is, arguably, most important. This is because the detection of these quality indicators is the first step in the decision process and is common to both the practitioner's and the examiner's decision. This is the case as they are both making their decision based upon the same mark which contains the same quality indicators. Their submission and usability thresholds, on the other hand, will be different. The

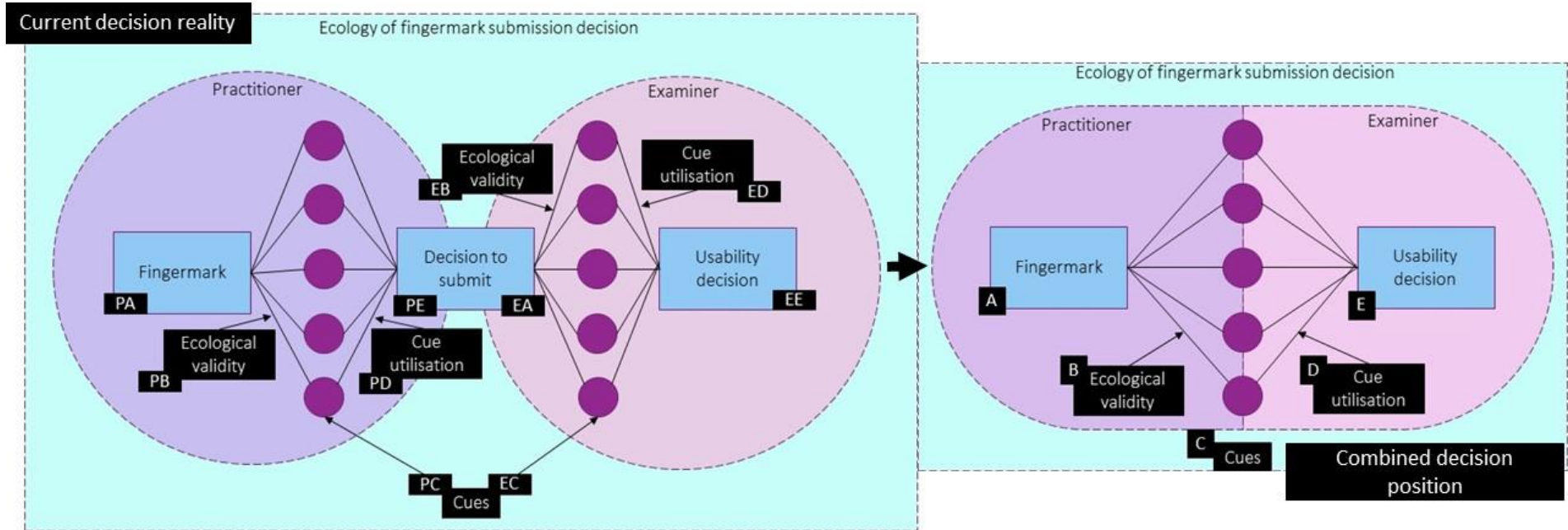
practitioner's submission threshold may be dependent on some or all of the factors mentioned above (crime type, case, or exhibit, for example), whereas the examiner's usability threshold is likely to be influenced by factors such as the task to be undertaken with the mark (for example identification or exclusion), whether there are suspects in the case for comparison purposes, or whether the mark is to be searched on a database such as Ident1.

A knowledge of the factors used during the decision making process (or quality signals) would provide a starting point for assessing similarities and differences in the important quality indicators used by the practitioners and examiners to make their respective submission and usability decisions. The application of Brunswik's lens model (Brunswik, 1952) as described by Hammond et al. (1964) can help as a framework against which to describe the fingerprint submission decision in terms of an attempt to mirror quality indicator cues between practitioners and examiners (Kahneman & Klein, 2009).

The traditional application of Brunswik's Lens Model is to model a judgement made in relation to a true state through the use of a number of weighted cues (Brunswik, 1952), for example (as described in 2.6.4), using weighted clues such as hair colour or clothing to determine the age of an unknown individual. If applying this idea to fingerprint submission decision it could be considered that the quality of the fingerprint is the true state to be assessed, and that the practitioner must correctly weigh up a series of cues indicating quality in order to ascertain this true value. Figure 4.2 illustrates two novel adaptations of this model that consider the fingerprint submission process (and associated quality judgements) at a more ecologically valid level.

Figure 4.2 depicts two adapted versions of Brunswik's Lens Model. The first version illustrates the current state of decision making within the fingerprint submission process. The fingerprint is the event criterion of the model which leads to a number of cues. These cues are the factors (or signals, in signal detection language) considered by the practitioner in making their decision to submit or discard. The value of each cue is determined and then weighted accordingly, and a final decision is reached. This leads to either a submitted or a discarded mark. Then a second decision process by which the examiner applies a set of weighted cues to this mark is modelled, resulting in the examiner's determination of usability. The second version of the model illustrates an idealistic view of the practitioner's submission decision process. In this version the process is modelled as a single judgement. The judgement being made in this case is not to determine a true solution or existing value, rather, it is to successfully determine (and mirror) the subjective judgement of another person (in this case the usability determination of a fingerprint examiner). This scenario differs from typical models and theories of group decision making (such as game theory) that describe competitive decision making (Robin, 1993), instead it places the emphasis on the practitioner decision as that

Figure 4.2 - An adaptation of Brunswik's Lens Model to illustrate the matching of the quality cue weights - Adapted from Newell and Shanks (2014)



- A** An event or criterion in the environment (in this case the fingerprint)
- B** Ecological validities – the relationship between the cues and the event/criterion
- C** The cues relied upon when making a decision
- D** Weighting and utilisation of cue information
- E** Making a judgement

which needs to utilise the appropriate cues and weights of these cues in order to make a submission decision appropriate to that of a fingerprint examiner. In this case the decision made by the practitioner relies upon similar cues and cue weights to that used during the examiner's usability decision. The addition of a situationally dependent moveable threshold means that the submission decision of the practitioner does not need to mirror the usability of the examiner in all cases. Instead the submission decision of the practitioner can be set at a lower threshold than that of the examiner to allow for difference in the identification and weighting of decision cues, and can also be lowered or raised according to changing environmental factors such as crime type and available resources.

In order to be able to apply the novel adaptation of Brunswik's Lens Model to the problem of fingerprint submission, knowledge of practitioner and examiner's decision cues needs to be gained, and decision cues common between practitioners and examiners need to be identified. Further exploration is then needed to establish the weight of each of these cues and to fully explore the similarities and differences between the two decision processes. A sound knowledge of the cues of both practitioners and examiners in version one of the model, will enable consideration of the ideal (standardised) content of the more efficient version 2 of the model.

Interviews with laboratory managers (summarised in 3.4.2. in relation Laboratory A, and Earwaker et al. (2015) in relation to MPS) highlighted a lack of policy or procedure in relation to fingerprint submission decision making. As mechanisms for mark submission are also not included within standard training packages (Lagden, 2014) it can be concluded that there is no standard, objective, procedure for making a fingerprint submission decision and that, therefore, this decision is highly subjective. Evidence for the subjectivity of this decision is provided within the individual differences data provided in Chapter 3 where a high degree of variation in decision making between practitioners was observed.

Previous research within the domains of crime and forensic science has suggested that there may be possible differences between the use of decision cues by examiners and practitioners. Baber and Butler (2012) compared the concurrent verbal protocols of novice and expert crime scene examiners in a scene processing task. They found that whilst novices reported considering the scene in terms of its key features, the experts were concerned with the likely actions that could be performed as a result of the examination (Baber & Butler, *ibid*), demonstrating a difference in approach. Garcia-Retamero and Dhamai (2009) investigated the use of decision cues when determining which of two properties were most likely to be burgled. Two expert groups, one consisting of both police officers and one of burglars, and a novice group, made up of students were asked to determine the cues likely to be important in deciding which property to burgle. The study found that the two expert groups differed

in the cues that they considered important in their decision making, with the cues determined by the police officer group found to be more similar to those of the novice group than to those in the other expert group (the burglars). This suggests that being an expert in carrying out a task (burglary in this case) may lead to the reporting of different decision cues in relation to that task, than those reported by an expert in attempting to understand how that task is carried out by others (as per the role of the police officers). The findings of these two studies suggest that there could be differences in the decision cues used by fingerprint practitioners and examiners. Examiners (who could be considered to be experts as, like the burglars studied by Garcia-Retamero and Dhamai (2009), they actually carry out the practical process of mark quality assessment that is required prior to carrying out a comparison) may consider cues more in terms of what could be done with the mark within the bureau such as exclusion, comparison, or search on an AFIS, given their experience in carrying out this task.

It can be concluded that fingerprint submission decisions are being made with different outcomes (as determined in Chapter 3 and Earwaker et al., 2015). The interesting extension of this finding is whether the decisions are being made by consideration of the same cues but that the practitioners are interpreting the cue information differently, or whether practitioners are using different cues to examiners. In the absence of a policy or procedure which dictates which cues should be used to dictate sufficiency for submission it may be that practitioners all employ similar cues in their quality judgement, or it may be that there are individual differences in approach.

Metacognition refers to the ability of an individual to have a critical awareness of their own cognitive processes (Flavell, 1979), and is commonly referred to as 'thinking about thinking'. This is a crucial consideration when attempting to determine rationale for fingerprint submission decision making, as practitioners need to be aware of the cognitive processes they are utilising during the decision making process in order to be able to communicate this cognition and, thus, provide accurate reasons for submitting or discarding a fingerprint.

In order to investigate the decision rationale used in fingerprint submission decision making it is beneficial to employ a qualitative research method in order to be able to fully explore the rationale provided by practitioners. Qualitative research methods allow the study of responses (verbal or written) through the application of one of many analytical methodologies in order to explore the meaning in these responses. A qualitative approach may involve some element of numerical analysis but will focus upon trends in the meaning of responses given, rather than to examine statistical correlations. As qualitative analysis draws meaning from the data itself, hypothesis are generated as a result of analysis, not prior to analysis (D. Silverman, 2011). Examples of qualitative data analysis methodologies include, content analysis, narrative analysis, and grounded theory. Content analysis

involves using pre-established categories and counting how many times these categories occur in a piece of text. Grounded theory, on the other hand, uses the text or responses to form the basis of a dynamic set of categories. This allows the data to inform the analysis and limits the potential for the loss of information. Narrative analysis looks to further examine the meaning in the narrative provided by a participant (i.e. how they tell a story as opposed to limiting analysis to the content of the story) (D. Silverman, *ibid*). Grounded theory appears the most appropriate of these methodologies to use in order to explore fingerprint quality assessment for a number of reasons. There has been no previous research assessing the rationale for the fingerprint submission decisions of laboratory practitioners. As such, pre-determined and fixed categories (as would be used in content analysis) would seem too prescriptive to allow open and comprehensive analysis in this novel area, whereas the grounded theory approach of establishing categories directly from the data and adding to these categories each time novel ideas are reported in the data, allows the consideration of all novel rationale for fingerprint submission. Grounded Theory is advantageous, in this case, over narrative analysis as the interest in this study lies in the submission decision rationale themselves, more than the method of communication of these rationale. As such, the present chapter takes a Grounded Theory approach to investigate fingerprint submission decision rationale, as further described in 4.3.3.

## 4.2. Summary of objectives

The current research seeks to identify the cues that are used by practitioners when determining whether or not to submit a mark, how these cues differ between practitioners, and how these cues compare to those used by examiners to determine usability in a mark. This will be achieved through addressing the following research objectives:

Objective 4.1. To explore the rationale behind laboratory practitioner fingerprint submission decision making within Laboratory A.

To explore the self-reported reasons for laboratory practitioner fingerprint submission decisions through completion of the following sub objectives:

Objective 4.1.1. A qualitative assessment of practitioner self-reported submission decision rationale within Laboratory A

To identify the self-reported decision rationale of practitioners.

Objective 4.1.2. An assessment of inter-practitioner variation in fingerprint submission decision rationale

To establish the level of variation in decision rationale between practitioners.



Objective 4.1.3. An assessment of practitioner fingerprint submission decision rationale according to experimental fingerprint

To establish variation in decision rationale according to fingerprint.

Objective 4.2. An investigation of inter-laboratory consistency in practitioner decision rationale

To explore consistency in laboratory practitioner fingerprint submission decisions between two UK fingerprint laboratories through completion of the following sub objectives:

Objective 4.2.1. Practitioner decision rationale within the MPS laboratory

To identify the self-reported decision rationale of practitioners.

Objective 4.2.2. Comparison of overall decision rationale between Laboratory A and MPS Laboratory

To compare the self-reported decision rationale of practitioners between the two laboratories.

Objective 4.2.3. A comparison of the effect of contextual information on practitioner decision rationale

To compare the effect of crime type information on the self-reported decision rationale between the two laboratories.

Objective 4.3. Exploring the relationship between practitioner submission decision rationale and fingerprint examiner usability rationale

To explore the relationship between practitioner submission decision rationale and examiner usability rationale through completion of the following sub objectives:

Objective 4.3.1. An assessment of fingerprint examiner usability rationale

To identify the self-reported decision rationale of practitioners.

Objective 4.3.2. A comparison of practitioner and examiner decision rationale

To compare the self-reported submission decision rationale of practitioners with the self-reported usability determination rationale of examiners.

## 4.3. Method

### 4.3.1. Overview of method

As part of the experimental research study discussed in Chapter 3 of this thesis, data was collected in relation to fingerprint examiners determinations of usability, and laboratory practitioner's reasons for

submitting or discarding a series of 30 experimental fingermarks. The rationale behind these practitioner and examiner decisions are explored according to the following method.

#### 4.3.2. Participants

Participants were fingerprint enhancement laboratory practitioners from two UK police force laboratories. 13 laboratory practitioners participated from Laboratory A and 11 laboratory practitioners participated from the Metropolitan Police Service Serious Crime Fingerprint Evidence Recovery Unit (MPS). These were the same practitioners who had participated in the study described in Chapter 3 of this thesis (see 3.3.2.2.) and constituted all practitioners employed in the two laboratories with the exception of those on long term leave or secondment. Four fingerprint examiners (two from each police force) also participated in this study. As data was collected simultaneously to participation in the study outlined in Chapter 3, the number of fingerprint examiners available to participate was limited due to the time consuming nature of the original fingerprint comparison task (set out in 3.3.2.1.). The additional independent fingerprint examiner who participated in the set up of the experimental data described in 3.3.2.1. was not included as a participant in this study to allow a direct comparison between practitioners and examiners working within the same organisations.

#### 4.3.3. Materials and procedure

##### Examiner self-reported rationale

Data in relation to fingerprint examiner usability decision making was gained during the completion of the mark assessment task carried out by fingerprint examiners as a preparative step in the compilation of a series of fingerprint images for use in the experimental work discussed in section 3.3.1.1.

Fingerprint examiners were provided with a series of 40 colour printed images of fingermarks developed with ninhydrin. Ten-print sets of the same source as the mark photographs were also produced (see 3.3.1.1 for further details). Source information was not, however, disclosed to the examiners. Examiners were asked to look at each fingerprint image and asked if they could put each back to source (i.e. state a match between the mark and a print). This provided a determination of the usability of the mark given a situation in which the examiner was equipped with the most complete possible fingerprint information (i.e. a good quality series of ten prints capturing all source fingerprint information). Examiners were asked to specify the usability, comparability and search-ability of each mark and were asked to provide justification for their determination (usability decision rationales).

#### Practitioner self-reported rationale

Following assessment of the 40 fingerprint images by fingerprint examiners (see section 3.3.1.1), two sets of 30 of these fingerprint images were collated (as utilised in Chapter 3). The first set was comprised of 10 borderline, or ambiguous, quality marks assessed by the majority of fingerprint examiners to be 'of sufficient quality for comparison', 10 borderline quality marks 'of insufficient quality for comparison' and 10 additional 'decoy' marks. The second set was comprised of the same 10 borderline sufficient and 10 borderline insufficient marks but included a different 10 'decoy' marks (see chapter 3.3.1.1) for further information relating to these marks). Both sets of 30 fingerprints were collated and bound within an A5 hard back sketch pad, with one image mounted per page. The first set of fingerprints was labelled as relating to a case of volume crime, the second set was labelled as relating to a serious crime. Practitioners were asked to look at each fingerprint image included in the image set in turn and to determine whether or not they would submit each of the fingerprints to the bureau in a casework situation. This task was carried out on two separate occasions by each practitioner. On one occasion the fingerprint pack provided stated that the 30 fingerprints provided related to a case of volume crime, on the other occasion the pack stated that the 30 fingerprints provided related to a case of serious crime.

Participants were also supplied with a response sheet. This sheet contained the necessary fields to input the data required for the completion of the experimental study outlined in Chapter 3 (whether or not the participant would submit each fingerprint, and how confident they felt in making this determination). In addition, the response sheet asked participants to 'please give any reasons for decision to mark-up/ not mark-up' for the purposes of data collection for the present experiment. Participants were provided with a free text box to provide a reason for their decision in relation to each of the 30 marks provided. The inclusion of a free text box ensured that practitioners were not directed towards a predetermined rationale, rather that they were required to communicate the rationale in the way they deemed to be most appropriate. This type of qualitative data collection was considered to be of paramount importance in this situation as there is a paucity of research in this area, and any suggested rationale that could have been included in a questionnaire type format to lead to a more quantitative analysis would have, in the absence of previous research, been heavily influenced by the experimenter's own view point, and may have not been representative of important decision factors. Equally the phrasing and breadth of rationale was of interest in this study, so it was essential that this information could be collected without the limitations of 'closed question' response fields.

#### 4.3.4. Analytical methodology

Data were analysed by employing a grounded theory style approach (D. Silverman, 2011). This approach employed a type of thematic coding to categorise the responses given, with the categories used stemming from the responses themselves (as opposed to categories pre-defined by the researcher). To do so, data were analysed by hand and categories (decision factors) were established in an iterative manner. Higher level analysis of the data was carried out through grouping the decision factors established into higher level categories based upon themes present in the data.

As participants had been provided with a free text field for the purpose of recording their decision rationale experimental data were qualitative and unbounded by nature. This qualitative aspect of the data was maintained through employing qualitative analytical methods, ensuring that the content, meaning, and any subtleties of the given responses were preserved throughout the analysis process, whilst allowing the results to be reported in a comprehensible style and analysed in line with the stated objectives.

Data were analysed in relation to Laboratory A, and the decision rationale were then compared between Laboratory A and the MPS Laboratory in order to identify any differences between the submission decision rationale between the two laboratories. Data were analysed according to laboratory and the datasets of the two laboratories were not combined. This ensured that individual differences in laboratory policy, procedure, and culture did not confound analysis, and also enabled exploration of differences in rationale within each laboratory as well as comparison of findings across the two laboratories. Data acquired from Laboratory A were analysed in the first instance. The order of analysis is important in this case as the grounded theory approach taken meant that categories of decision (termed as decision factors) were added as required by the data, as such, a considerable number of decision factors were already in existence upon consideration of the second data set (MPS data), new factors, were, again, added as required, and any factors not utilised in the second set of response were removed from the results set.

For each laboratory the responses of each practitioner in relation to each of the 20 'borderline' experimental fingermarks presented were recorded. A system of categorising responses was developed which was based upon the qualitative analysis tool Grounded Theory (D. Silverman, 2011). Responses were recorded in the order of the participant and mark in consideration (ie. beginning with participant A, mark A, then participant A, mark B etc). Each novel decision reason given was recorded as a new category heading (termed a factor), and any reason given which fitted into an existing category was recorded as such. This fluid approach to categorising decision rationale meant the data could be categorised at a workable and intelligible level without losing valuable information, as may have been the case had categories been predetermined. Within the responses provided within the

course of experimental work outlined in Chapter 3 practitioners had determined whether or not they would submit each of the experimental mark to the fingerprint bureau, as such the data set included rationale relating to a combination of self-reported reasons for submitting and for discarding fingerprints. Indeed, some reasons for submission included rationale for submitting and discarding a fingerprint; for example: 'ridge flow is broken up, but pattern is present' provided as a reason for submitting a mark. In this example 'pattern present' would be a positive rationale (or a positive quality indicator), whereas 'broken up ridge flow' would be a negative rationale (or a negative quality indicator). The grounded theory style analysis employed allowed the consideration of multiple decision rationale according to key decision themes. For example, 'ridge flow is broken up, but pattern is present' would be recorded as two decision rationale, one relating to broken ridge flow, and one relating to the presence of a pattern. As such, the data represented all aspects of cue consideration in relation to a mark as opposed to just those indicating a submission decision. This was intended to allow a full consideration of cues involved in the fingerprint quality assessment process, regardless of the ultimate outcome of that decision.

The factors established were then analysed at a higher level in order to draw out broader trends. A thematic analysis was employed and factors were grouped into categories according to the type of fingerprint quality indicators they related to. In keeping with the grounded theory approach categories were based upon the decision factors within the data, not pre-determined by the researcher, and data were hand coded.

#### 4.4. Results

Rationale data are included within Appendix C.1.

##### 4.4.1. Objective 4.1: Analysis of metacognition within Laboratory A

###### 4.4.1.1. Objective 4.1.1. – Overall decision rationale Laboratory A

###### Summary of data

Data gathered from the Laboratory A provided a total of 780 decisions made by 13 practitioners each making 60 submission decisions. Practitioners provided, on average, 1.4 decision factors for each decision made; giving an overall total of 1092 decision rationale.

###### Identifying and illustrating decision factors

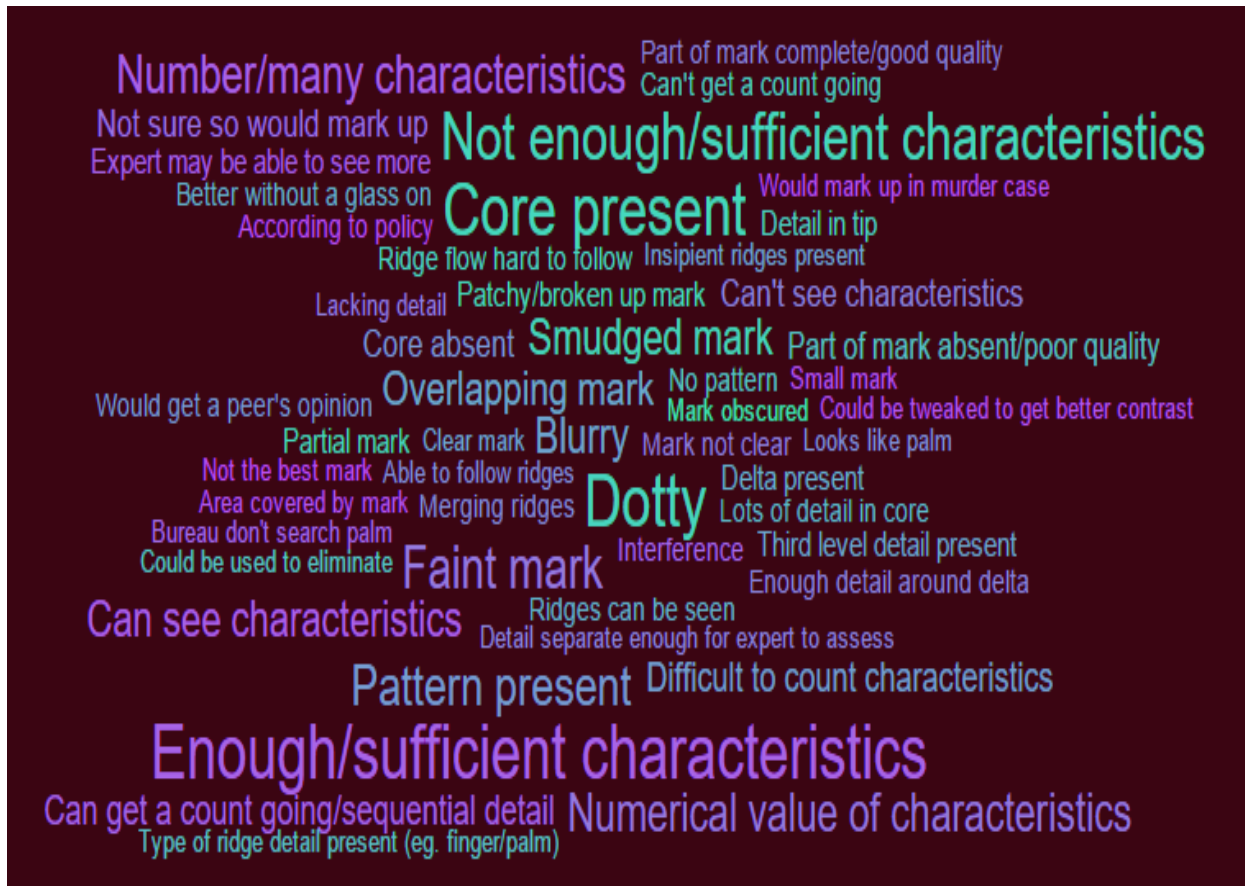
Through initial grounded theory style analysis, responses were categorised in to a total of 52 decision factors (with the addition of a 'no response' category for the 172 cases in which practitioners had chosen to not provide a decision reason).

The frequency of the 52 decision factors was illustrated using a word cloud (see Figure 4.3). The word cloud was produced using <http://worditout.com/>, a freely available online programme which identifies the frequency of words in a collection of text provided and produces a graphic illustration representing the frequency of the word as its relative size in the illustration. Figure 4.3, therefore demonstrates each of the 52 decision factors provided and also provides an initial illustration of the prevalence of these decision factors within the data (coloured text within the word cloud was randomly generated and is not indicative of data grouping). With the exception of 'no response' (provided on a total of 172 occasions, and not included within the word cloud generated), the most commonly stated decision factor provided was 'not enough/sufficient characteristics' which was stated a total of 85 times. Other common responses were 'core present' (84 responses), 'dotty' mark (81 responses), and 'not enough/sufficient characteristics' (71 responses). Examples of factors only recorded once during analysis were 'insipient ridges present', 'lacking detail', and 'small mark'. This data suggests sufficiency of characteristics and the presence of a core to be important factors, along with the clarity of the ridge detail present.

Displaying the decision factors in this way allows easy observation of the differences in frequency whilst still allowing clear observation of all decision factors provided. Numerical instances of the all decision factors are provided in Table 4.1. in relation to the categorisation of decision rationale. Interesting, the data set includes a number of less frequently occurring decision factors which suggest a consideration of the perspective and processes of a fingerprint examiner during the practitioner's submission decision process. Examples include 'expert may be able to see more', 'Bureau don't search palm', and 'could be used to eliminate'.

The existence of 52 different decision factors and the range of the meaning of these factors suggests inconsistencies within the laboratory in decision making rationale and demonstrates that there is not one key factor, or a small number of key factors, that are being consistently being employed; rather that a large and diverse range of factors are being used.

Figure 4.3 - Word cloud illustrating all decision factors provided by Laboratory A practitioners, and the relative frequency of these factors



#### Categorising decision factors

The recorded decision factors were then grouped into 9 decision categories. The decision categories were determined through consideration of common themes within the 52 decision factors provided (thematic analysis with themes established by hand, as opposed to computer software, based upon the data present). Higher level categorisation of the decision factors was deemed necessary to be able to draw out higher level themes and patterns within the data. The decision categories identified were those relating to:

- Clarity of the mark
- Pattern (1<sup>st</sup> level detail)
- Characteristics (2<sup>nd</sup> level detail)
- The area and type of mark present
- Policy or situational factors
- The presentation of the fingermark
- Comparative quality statements
- Unsure

- No response given

Table 4.1 illustrates the inclusion of the 52 decision factors within the 9 decision categories.

Explanation of the content and meaning of decision categories defined

#### *Clarity of the mark*

Fifteen decision factors were determined to sit within a decision category of 'clarity of the mark'. This category included any decision factor which referred to the ease with which the practitioner could determine the presence of the ridge detail within the enhanced fingerprint. This included factors such as the level of contrast between the mark and the background surface and the presence or absence of continuous ridge flow. A variety of phrases were used to allude to mark clarity. Examples include dotted mark, broken up mark, and smudged mark. Interestingly the majority of decision factors provided in relation to clarity suggested negative quality indicators. This may suggest that mark clarity is a starting point in the quality determination process; if the mark is not of sufficient clarity that the detail contained within it can be observed then this is likely to be the end point of the decision process, however should the mark be of good quality it may be that a more advanced factor becomes that reason for the decision made.

#### *Pattern, or 1<sup>st</sup> level detail*

Five decision factors fitted within the decision category of fingerprint pattern. These included the presence or absence of a pattern, core, and delta. These factors were a mixture of positive and negative decision reasons, suggesting that the absence of a pattern or core and delta features can be a reason to decide not to submit a fingerprint, but equally that the presence of such features can form the basis of a decision to submit a mark.



Table 4.1- The categorisation and occurrence of Laboratory A decision factors

Clarity of mark	Freq.	Pattern/1st level detail	Freq.	Characteristics/ 2nd level detail	Freq.	Area and type of mark present	Freq.	Policy/situational factors	Freq.
blurry	38	pattern present	53	enough/suff characteristics	85	partial mark	8	according to policy	6
dotty	81	core present	84	not enough/suff characteristics	71	area covered by mark	1	FPB don't search palm	1
overlapping mark	35	core absent	19	numerical value of characteristics	43	part of mark complete/good quality	6	would mark up in murder case	1
faint mark	56	no pattern	7	number/many characteristics	49	part of mark absent/ poor quality	13	could be used to eliminate	1
smudged mark	40	delta present	8	difficult to count characteristics	23	type of ridge detail present (finger/palm etc)	1	Total frequency of all factors	9
patchy/broken up mark	8	Total frequency of all factors	171	can't see characteristics	17	small mark	1		
ridges can be seen	4			can see characteristics	38	looks like palm	2		
able to follow ridges	1			can get a count going/sequential detail	21	detail in tip	9		
ridge flow hard to follow	5			lots of detail in core	6	Total frequency of all factors	41		
detail separate enough for expert to assess	2			enough detail around delta	6				
merging ridges	5			insipient ridges present	1				
mark not clear	4			third level detail present	9				
interference	4			lacking detail	1				
mark obscured	2			can't get a count going	5				
clear mark	2			Total frequency of all factors	375				
Total frequency of all factors	287								

Presentation may make a difference	Freq.	Comparative quality statement	Freq.	Unsure	Freq.	No response given	Freq.
better without a glass on	5	not the best mark	2	would get a peer's opinion	4	no response	172
could be tweaked to get better contrast	3	Total frequency of all factors	2	not sure so would mark up	20	Total frequency of all factors	172
expert may be able to see more	10			Total frequency of all factors	24		
Total frequency of all factors	18						

### *Characteristics or 2<sup>nd</sup> level detail*

There were a total of 14 decision factors which fitted within the decision category of characteristics, or 2<sup>nd</sup> level detail. This category included any factors which related to the presence or absence of characteristics within the mark as well as factors which related to the numerical value or sufficiency of these characteristics. Some factors included reference to detail within the core or delta of a fingerprint and so were included within this decision category as they referred specifically to the detail within the pattern or feature as opposed to the presence or absence of the pattern or feature. Again there was a mixture of decision factors provided to explain submission decisions and also decisions made not to submit a mark. The decision factors 'can get a count going/sequential detail' and 'can't get a count going' seem to allude to an acknowledgement during the practitioner decision making process of the need for consecutive minutiae to be present in a mark in order for a fingerprint examiner to be able to make an identification.

### *Area and type of mark present*

The area or type of mark present contains eight different decision factors. These factors allude either to the physical area covered by the mark, the completeness of the mark, or the type of area of the mark present (for example, palm or fingertip). Again, a mixture of positive and negative decision factors are included within the category.

### *Policy/situational factors*

This decision category contains the four decision factors given that included reference to policies, procedures, or processes. These are reasons suggestive of a decision being made as a result of the acknowledgment of processes, procedures or workflow agreed upon by the organisation. The majority of these factors do not acknowledge the aspect of the mark that leads to a policy or procedurally driven decision; the exception to this being '[the fingerprint bureau] don't search palm' which provides the information that because the mark is considered by the practitioner to be palm they will use the policy that palm cannot be searched in order to inform their submission decision.

### *Presentation may make a difference*

The three decision factors included within the 'presentation may make a difference' decision category, relate to aspects of mark enhancement and differences between the working practices of the laboratory and the bureau that may mean a change in the quality, or perceived quality, of the mark. All three decision factors are of interest when considering the effectiveness of the mark filtering process and the transition of ridge detail from the laboratory to the bureau. 'Better without a glass on' refers to differences in perceived quality of a mark according to whether or not an eye glass is

used to magnify the detail. When working on paper copies of marks it is common for fingerprint examiners to make use of such a magnifying tool in order to aid in their comparison, practitioners may use such an eye glass to aid in mark quality assessment but may also make the decision without the use of an eye glass. 'Could be tweaked to get better contrast' alludes to the imaging process that sits between the practitioner's quality assessment and the examiners analysis comparison. The practitioner will commonly make their submission decision based upon the actual ridge detail on the exhibit. Ridge detail deemed to be sufficient will then be photographed or imaged and the image may then be tweaked to improve contrast or converted to black and white. It is the resultant image that is viewed, either in paper form or electronically by the examiner. The final decision factor included within this category simply acknowledges that 'an expert may be able to see more in the mark'. This may be an acknowledgement of differing expertise between the practitioner and examiner groups. These decision factors hint at the acknowledgement of an importance of adjusting a submission decision threshold to take into account these differences in mark presentation.

#### *Comparative quality statement*

One decision factor was found to be ambiguous in nature. In order to avoid making assumptions about the data, which may incorrectly attribute meaning to the rationale and assign it to the incorrect category, the data was included within its own decision category. The decision factor 'not the best mark' did not fit neatly within any of the previously described decision categories as it was unclear whether the practitioner had meant that the mark was not good quality in direct comparison to the other marks observed during the exercise, or whether it was simply a slang term referring to the mark not being of a high quality. Such an ambiguity exposes a limitation of data collection in a questionnaire format. Had this data been gained through interview the precise meaning of this rationale could have been explored.

#### *Unsure*

Two decision factors were provided that suggested that a decision had been made based upon indecision. The decision factors suggested either that the practitioner would ask the opinion of a peer and then make a decision or would err on the side of caution and submit the mark if unsure. These are both interesting rationale which provide insight into the decision making process. Stating that if you were unsure you would seek the opinion of a peer highlights that there is a lack of useful feedback within the submission decision process as, instead of asking for an examiners opinion, the practitioner seeks the opinion of a fellow practitioner, and does not then gain an idea of quality determination from an examiner's view point. Submitting the mark if unsure suggests that the practitioner may place a higher utility value on submitting a poor quality mark than on discarding a good quality mark, and

so has taken this more cautious approach and has lowered the threshold for submission when the appearance of the mark has made the decision uncertain.

#### *No response given*

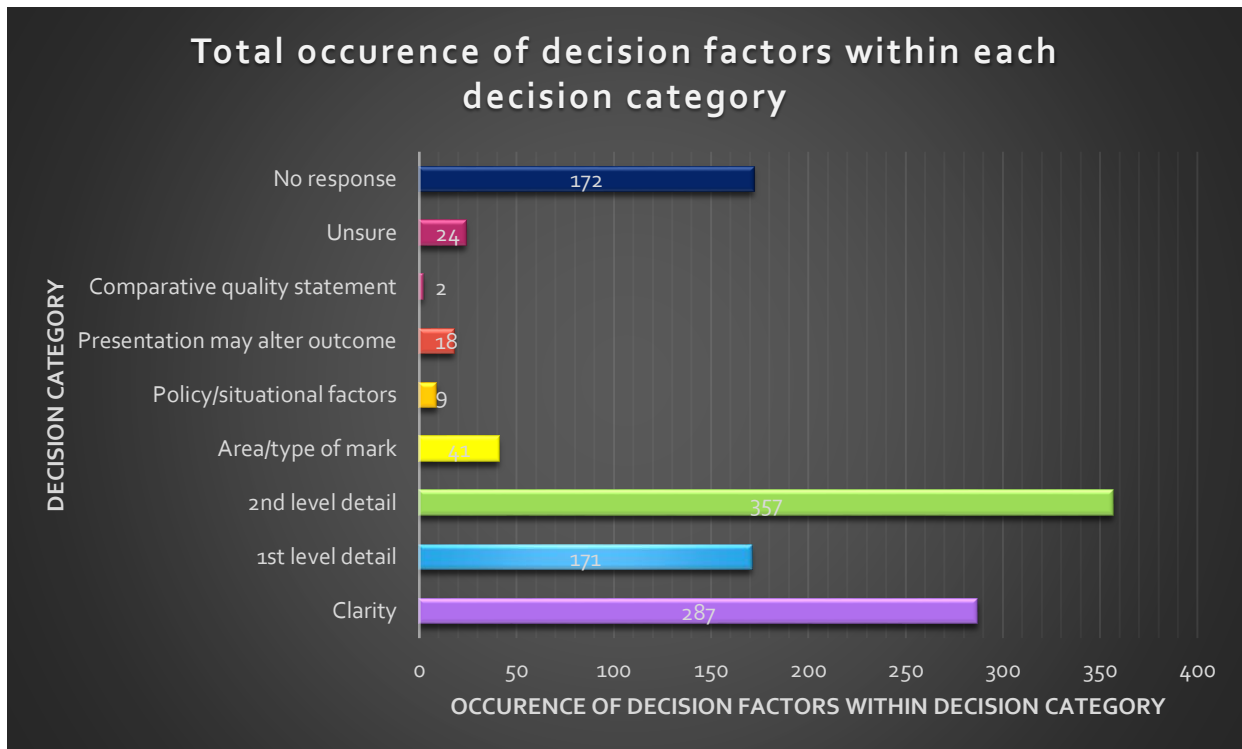
In some cases, practitioners did not complete the rationale box provided within the response form they were given. This may have been because they considered this to be unnecessary information, additional to the experimental task, or did not want to spend the additional time on the task. One practitioner did not respond in relation to any of the decisions that they made on either of the two occasions that they participated in the study. Other participants responded on one participation occasion but not on the other. Others provided rationale in relation to some of the decisions that they made but declined to do so in relation to others. This may have been because they felt that the task was overly repetitive or may have been that they found it challenging to describe their rationale in certain cases, or may have been because they did not feel it was necessary to describe their rationale in non-challenging cases. The assessment of correlation between 'no response' rationale and decision performance may be an interesting form of further analysis in order to be able to glean further information from this data.

The categorisation of decision factors enables a more quantitative discussion of decision rationale at this higher level.

The frequency total for each decision category provided in Table 4.1, shows 2<sup>nd</sup> level detail to be the most frequently occurring decision category, followed by clarity of the mark. Lesser occurring categories were comparative quality statements, policy or situational factors and the idea that presentation may alter the decision outcome.

To further explore and quantify the total proportion of decision categories these were presented in a bar chart (see Figure 4.4).

Figure 4.4 - Bar chart illustrating the total occurrence of decision factors within Laboratory A

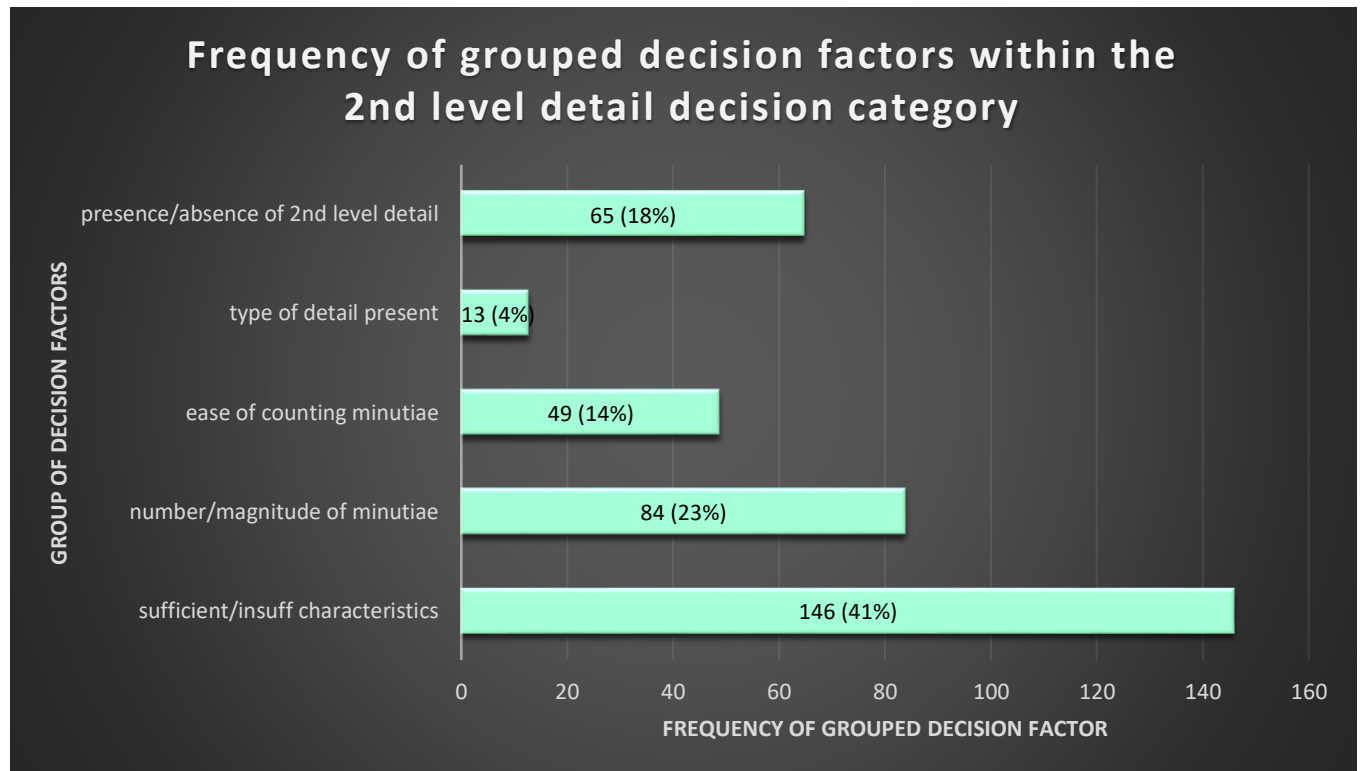


Further exploration of 2nd level detail as a common decision rationale

Division of the Laboratory A decision rationale data according to defined decision categories demonstrated the category of 2<sup>nd</sup> level detail to be the most commonly reported. As such, this category was selected to be further analysed at a lower level in an attempt to better understand the aspects of 2<sup>nd</sup> level detail that were being used during the fingerprint submission decision process. Figure 4.5 illustrates five subcategories within the 2<sup>nd</sup> level detail category and the frequency of responses provided within each of these sub categories. Sub categories were established through consideration of the decision factors included within the category. These were grouped into subcategories according to similarities in meaning. This avoided the repetition of similar data within different categories; such as in the case of positive and negative versions of the same rationale, for example 'sufficient characteristics present' and 'insufficient characteristics present'. Similarly, all references to numerical values of minutiae were grouped together. Grouping of the data in this way led to the creation of the following subcategories within the 2<sup>nd</sup> level detail category:

- Presence/absence of 2<sup>nd</sup> level detail
- Type of detail present
- Ease of counting minutiae
- Number/magnitude of minutiae
- Sufficient or insufficient characteristics

Figure 4.5 - The composition of decision factors within the 2nd level detail category



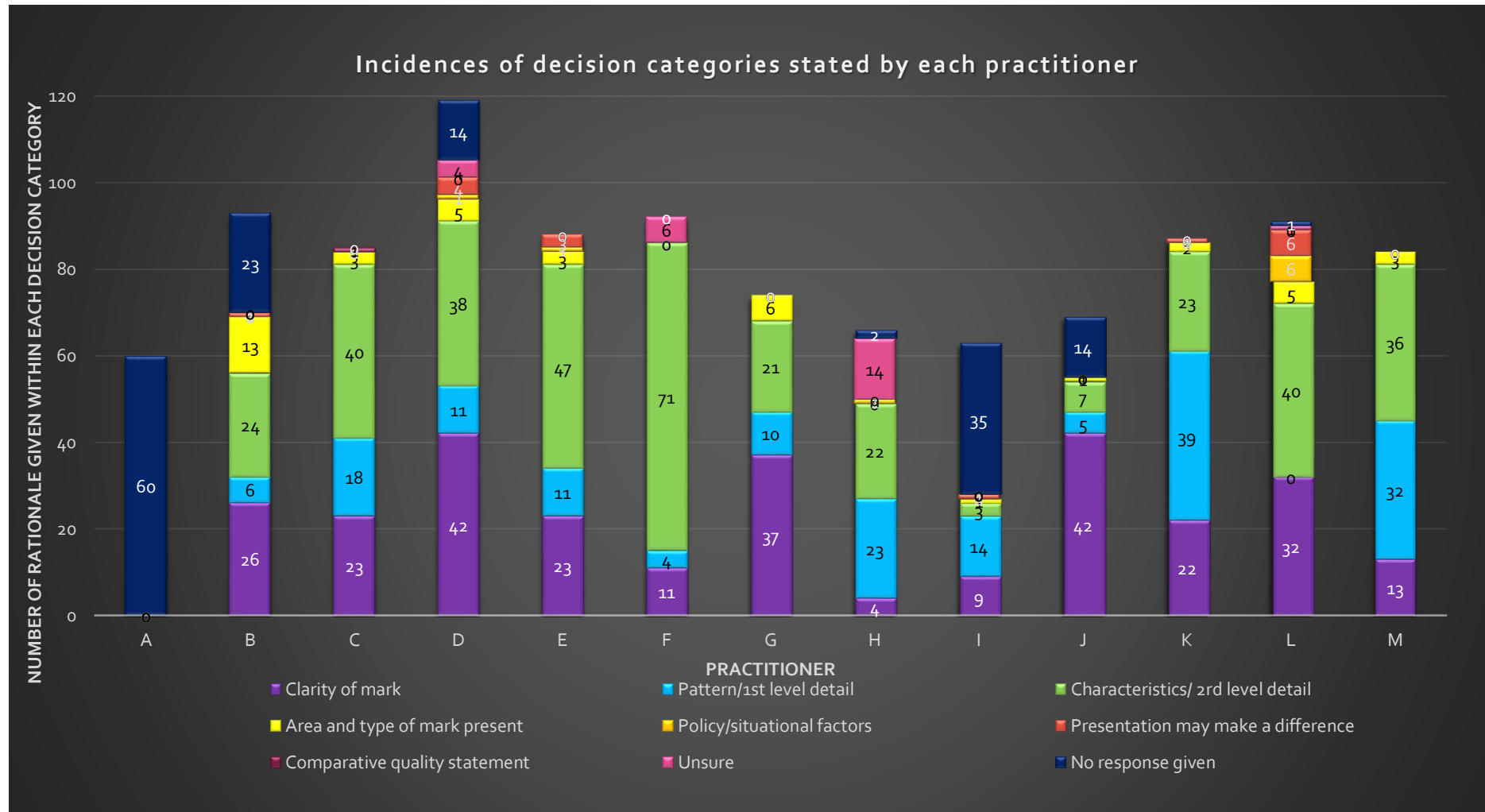
The data presented in Figure 4.5. shows the sufficiency or insufficiency of characteristics to be the largest contributor to the 2<sup>nd</sup> level detail category. Factors within this sub category were provided on 146 occasions. The second largest sub category contained factors relating to the number or the magnitude of minutiae present, provided on 84 occasions. Overall factors which indicated that a count or threshold of minutiae was important (sufficient/insufficient characteristics, number/magnitude of characteristics and ease of counting minutiae) made up a total of 78% of the responses within this category.

#### 4.4.1.2. Objective 4.1.2. Decision categories according to practitioner

The sum of decision factors provided within each decision category were compared for each of the 13 practitioners from the Laboratory A who participated in the study. The make-up of decision categories for each practitioner (A – M) is shown in Figure 4.6.

Figure 4.6 illustrates the variation between participants in the number and category of rationale provided during the experimental exercises. The degree of inter-practitioner variation in decision categories stated suggests considerable individual differences in self-reported decision rationale. However, for the majority of practitioners 2<sup>nd</sup> level detail was the most commonly cited reason for making a decision. The majority of practitioners provided more than one reason for some of the decisions that they made. Participant D gave the highest number of decision factors per decision

Figure 4.6 - Incidences of decision categories according to Laboratory A practitioner



made. Participant A gave the lowest number of reasons by not responding to any of the rationale questionnaires during either of the experimental exercises. Incidences of 'no response' were not spread equally amongst the remaining 12 participants; 6 further practitioners provided some 'no response' decision factors during the course of the two exercises. Intra participant variation shows that all practitioners use more than one decision category to justify their decision making across the marks provided. No practitioner reported that they solely rely upon one category or type of information in the marks to make their decisions.

#### 4.4.1.3. Objective 4.1.3. – Decision rationale according to experimental fingermark

##### Decision categories according to mark

The findings detailed within Chapter 3 and Earwaker et al. (2015) demonstrated differences in decision success according to experimental fingermark. Data in relation to decision rationale for all decisions made by all practitioners from laboratory A was divided according to experimental fingermark for the 20 borderline marks examined within Chapter 3 (a total of 520 decisions).

Figure 4.7 - Make up of decision factors according to experimental fingermark

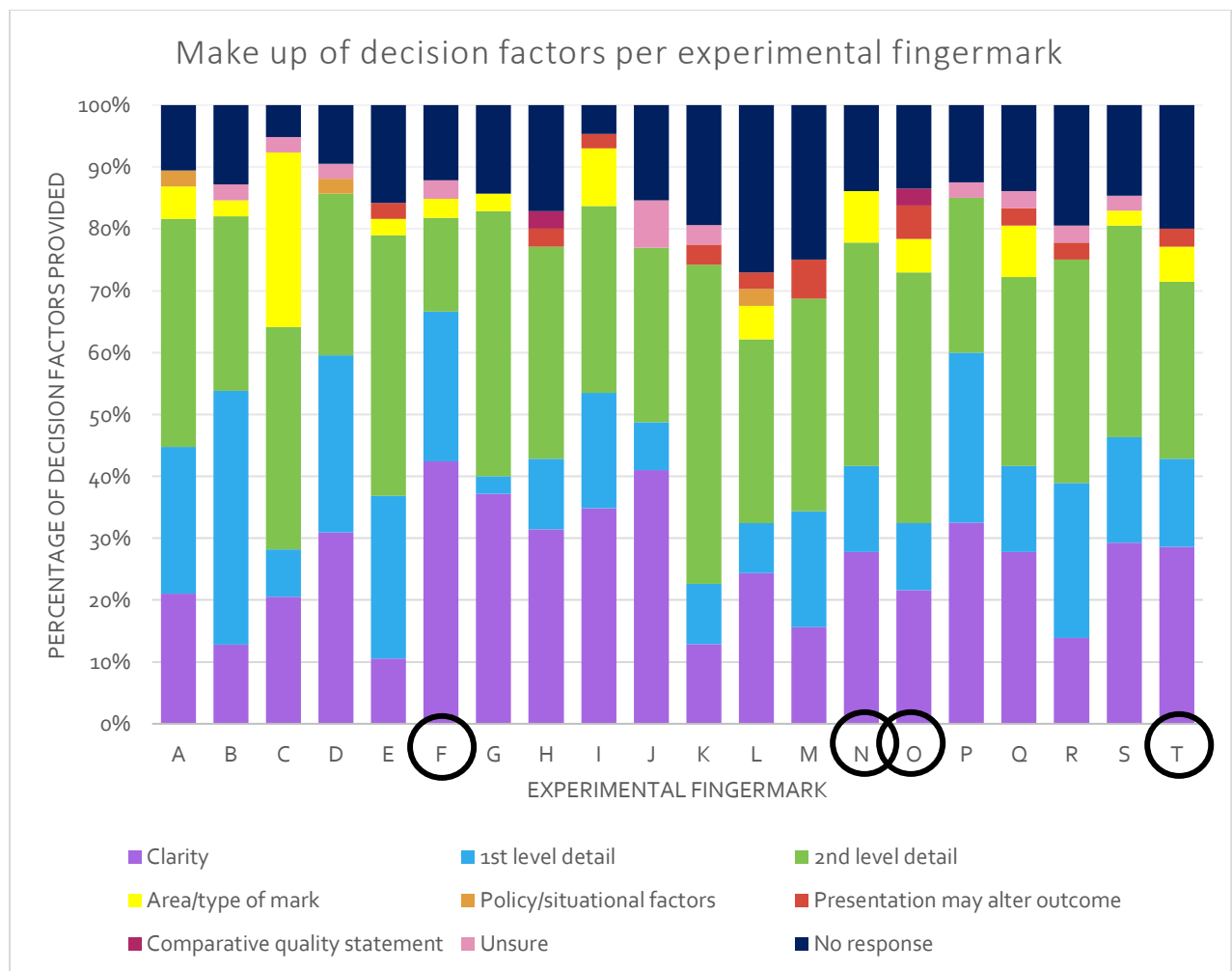




Figure 4.7 demonstrates similarities in the decision factors reported in relation of each of the experimental fingerprints. Clarity, 1<sup>st</sup> level detail, and 2<sup>nd</sup> level detail decision factors were reported in relation to all fingerprints. There are however, slight differences in the proportion of decision factors stated for each mark, for example a relatively small proportion of 1<sup>st</sup> level detail decision factors reported in relation to mark G and a relatively large proportion of area/type of mark decision factors reported in relation to fingerprint C. Fingerprints F, N, O, and T have been highlighted within Figure 4.7. These were the four marks highlighted in Chapter 3 as having the highest practitioner submission error in relation to examiner usability determinations. Consideration of the make-up of decision categories for these four fingerprints does not reveal a pattern in rationale for submission decision making in relation to these marks. Mark F can be seen to have the highest proportion of rationale within the 'clarity' category of all marks, but clarity rationale are not particularly high for the remaining three marks.

#### 4.4.2. Objective 4.2: Inter-laboratory consistency in decision rationale

##### 4.4.2.1. Objective 4.2.1. Overall decision rationale within the Metropolitan Police Serious Crime Fingerprint Enhancement Laboratory

###### Summary of data

Decision rationale data gathered from the MPS Serious Crime Fingerprint Enhancement Laboratory related to a total of 660 decisions made by 11 practitioners in relation to the consideration of 60 fingerprints. Practitioners provided a total of 684 decision rationale; an average of 1.04 rationale given per decision made.

###### Determining decision factors

As with responses provided by Laboratory A, decision rationale data for the MPS lab was initially considered through the use of Grounded Theory style analysis as described in the method (section 4.3.3). As the MPS data was analysed after the initial analysis of the Laboratory A data a number of decision factors had already been determined. These decision factors were used as a starting point for the Grounded Theory style analysis. As such, any of the decision rationale which fitted within the pre-existing decision factors derived from the Laboratory A data were recorded as such but, in keeping with the principles of Grounded Theory, any responses which were not a good fit for the existing factors led to the creation of new decision factors. Through this initial Grounded Theory style analysis, responses within the MPS data were categorised in to a total of 77 decision factors (with the addition of a 'no response' category for cases in which practitioners had chosen to not provide a decision reason). Of these decision factors 36 were brought forward from the Laboratory A analysis (of the total of 52), and an additional 41 factors identified within the MPS data.

The frequency of the 77 decision factors was illustrated using a 'word cloud' (see Figure 4.8). The word cloud was produced as outlined in section 4.4.1.1. This produced a graphic illustration representing the frequency of each decision factor as its relative size in the illustration. Figure 4.8, therefore, demonstrates each of the 77 decision factors provided within the MPS responses and Table 4.2 also provides an initial illustration of the prevalence of these decision factors within the data. The frequency of each decision factor is also provided in.

With the exception of 'no response' (provided on a total of 243 occasions, and not included within the word cloud generated), the most commonly stated decision factor provided was 'not enough/sufficient characteristics' which was stated a total of 102 times, followed by 'pattern present' with 95 responses. Responses given moderately frequently were stating a 'numerical value of characteristics' (42 responses), stating that the 'number or many characteristics' were the reason for decision making (58 responses), or 'can see characteristics (no mention of quantity)' (58 responses), 'overlapping mark' (20 responses), 'core present' (38 responses). Examples of factors only recorded once during analysis include 'open to interpretation', 'insufficient', and 'gut instinct'. This data suggests sufficiency of characteristics and the presence of a pattern to be important factors, and alludes to the number of characteristics present being an important determining factor.

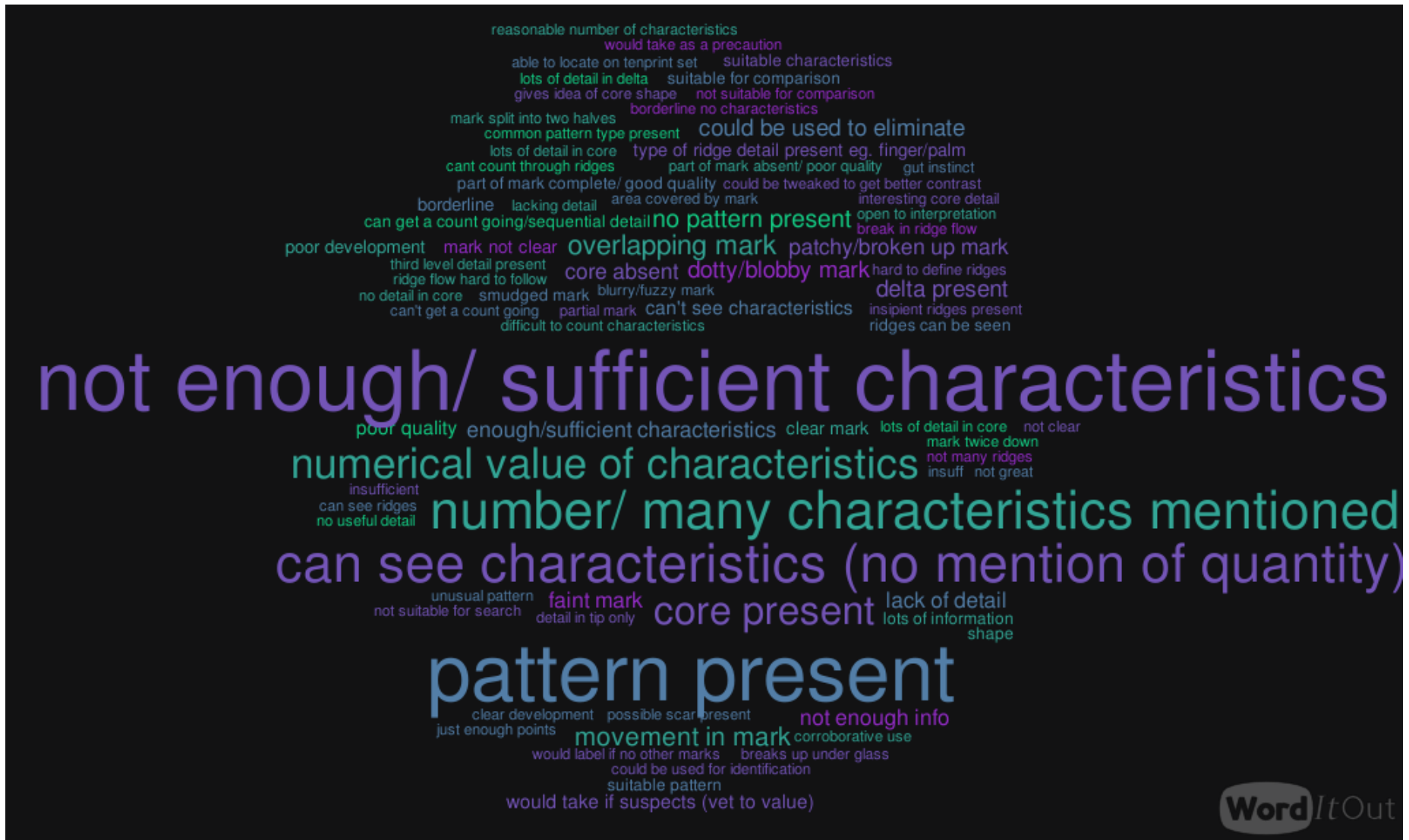
Displaying the decision factors in this way allows easy observation of the differences in frequency whilst still allowing clear observation of all decision factors provided. The data collected from the MPS laboratory demonstrates a considerable diversity of decision factors. The majority of these factors have only been provided on one occasion. There are, however, themes amongst the decision rationale provided which were further explored through the categorisation of decision factors (described below).

#### Determining decision categories

In order to take a higher level approach to the consideration of the MPS data, decision factors were grouped into a number of decision categories. As a starting point for this analysis, the existing decision categories identified during analysis of the Laboratory A data were utilised and decision factors from the MPS data were assigned to the appropriate decision category. Using the same decision categories was deemed to be a valid approach to categorising the MPS data as the existing categories appeared to be a good fit for the new data, and the use of this approach would allow easily comprehensible comparison of the two data sets. As such, decision factors were assigned to the nine existing categories as illustrated in Table 4.2.

One decision factor failed to sit neatly within any of the predefined categories and, resultantly, was categorised separately. The composition of decision factors within each category is discussed below.

Figure 4.8 - 'Word cloud' displaying MPS decision factors and their relative frequency within the data





Description of decision categories

*Clarity factors*

A total of 18 decision factors relating to the clarity of the developed ridge detail are included in this decision category. The most common situations in which clarity was specified as a decision reason were situations in which the mark was overlapping (20 occurrences) and instances of movement in the mark (17 instances).

*Pattern/1st level detail factors*

Nine factors that suggest pattern or 1<sup>st</sup> level detail to be important in submission decision making were provided within the MPS data. These include the presence or absence of a pattern, and the presence of a core or delta within that pattern. 'Pattern present' is the most common decision factor within this category, with a total of 95 occurrences in the data. Decision factors within this category also include ideas relating to the perceived usefulness of the pattern type present. For example, a 'suitable pattern', or commenting on the occurrence of the pattern type within the population ('unusual pattern' and 'common pattern type present').

*Characteristics/2<sup>nd</sup> level detail factors*

Characteristics/2<sup>nd</sup> level detail factors is the largest decision category, containing 22 different decision factors with a combined occurrence of 308. The most common decision factor within the MPS data is 'not enough/sufficient characteristics' with 102 occurrences. The quantity of characteristics and specifying a numerical count of characteristics as a decision reason are also common decision factors.

*Area and type of mark present factors*

One additional decision factor was provided within this category within the MPS data. 'Possible scar present' was mentioned as a decision factor and as this would have an impact in the type of mark present and also the area of ridge detail it was included within this decision category.

*Policy and situational factors*

A marked difference could be seen between the decision factors provided by MPS and Laboratory A in relation to policy and situational factors. A number of the situational factors provided by MPS focus upon the ultimate use of the mark as the determining factor in submission. Rationale such as 'sufficient/ insufficient for comparison', 'could be used to eliminate' and 'suitable for identification' suggest a consideration of differing uses of the ridge detail by examiners. The inclusion of such reasoning would suggest an understanding of the differing requirements for each of these processes. In order for this to be an effective method of mark discrimination, information as to which of these

processes was to be used by an examiner would need to be communicated to the practitioner in each case.

#### *Presentation may make a difference*

Only one decision factor, stated once within the MPS data, fell into the decision category of 'presentation may make a difference'. This was the rationale that a mark could be 'tweaked to get better contrast'. This reasoning acknowledges the potential for there to be a difference between the quality of a mark as viewed in situ on the exhibit by the practitioner, and as viewed as an electronic or printed image by the examiner.

#### *Comparative/general quality statement (threshold) factors*

This decision category was created in order to house a decision factor within the Laboratory A data that did not easily sit within any other category. Within the MPS data were six decision factors which fitted with the idea of a comparative quality statement. The original factor included within the Laboratory A data ('not the best mark') had been ambiguous as to its meaning, whereas the decision categories included within this category from the MPS data fit the category as they are comparative statements, some of which allude to a comparison against an internal threshold, or statements which generally express the level of quality in a mark in a subjective manner. As such the title of the category has been expanded to be more representative of the factors included within it. 'Not enough info' was one of the decision factors included within this category. This factor was assigned to the current category as opposed to the 'unsure' category as it was felt that the factor was relating to information within the mark as opposed to extraneous information needed in order to make a decision. Had this been the case then this factor may have lent itself more to inclusion within the 'unsure' category. Equally, 'lots of information' was also included within this category.

#### *Unsure factors*

Only one decision factor was found to fit within this decision category. 'Would take as a precaution' was stated on one occasion by one practitioner as rationale for a decision to submit a mark. This statement hints at the perceived importance of not missing a mark and also hints at the possibility that the practitioner places a higher utility value on sending a poor quality mark than discarding a good quality mark.

#### *Instinct Factors*

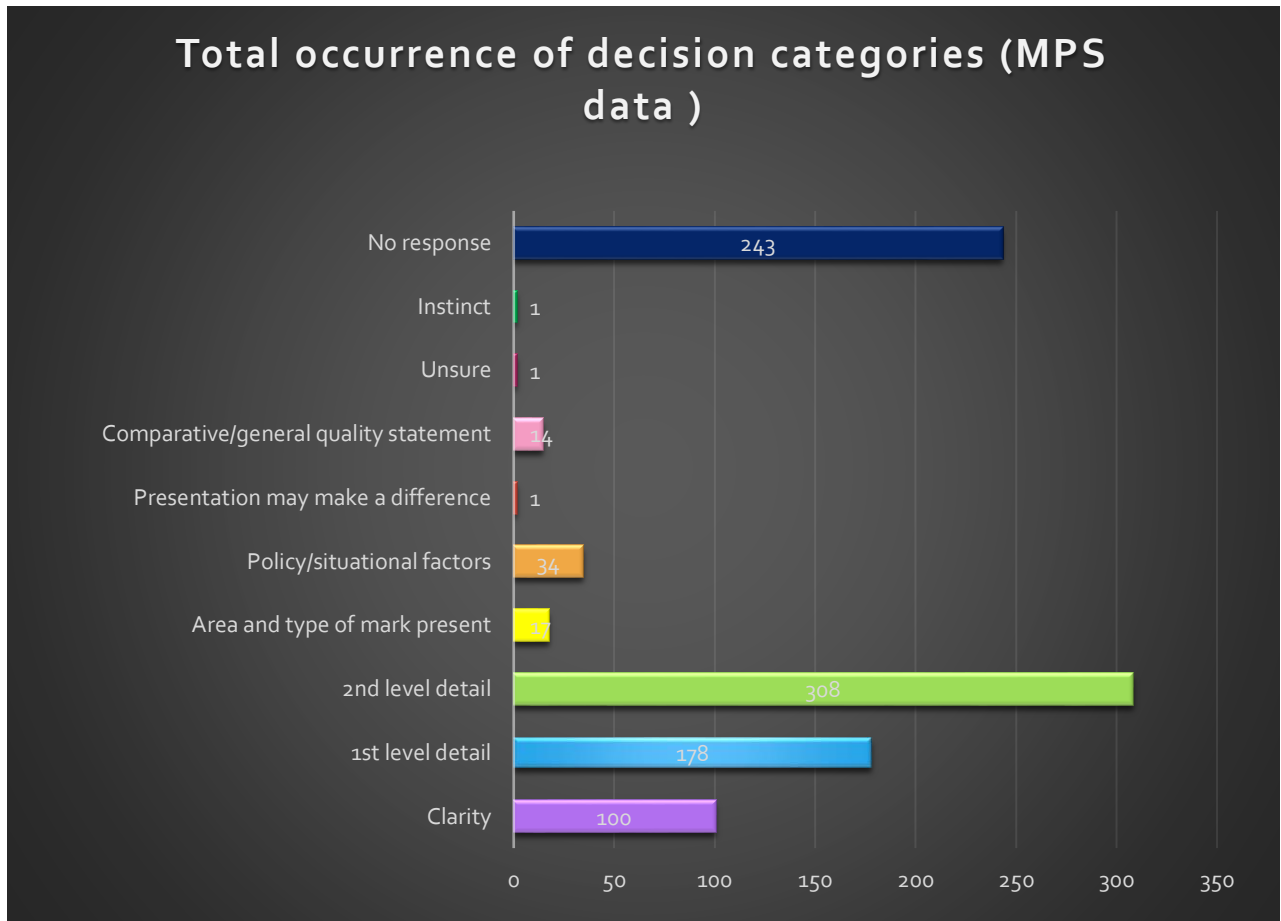
The decision factor 'gut instinct' did not fit into any of the pre-existing decision categories defined during the analysis of the Laboratory A data. As a result, the category of 'instinct factors' was created for this factor alone.

No response

In 243 of the decisions made during the experiment no decision rationale was provided by the practitioner.

Figure 4.9 illustrates the number of decision factors provided by MPS laboratory practitioners within each of the decision categories.

Figure 4.9 - Occurrence of decision categories within the MPS Laboratory



2<sup>nd</sup> level detail can, indeed be seen in Figure 4.9 to be the category containing the highest number of responses (308), followed by first level detail (178) and clarity (100). The number of no responses is reasonably high at 243. Only 1 example of 'unsure' 'presentation may make a difference' and 'Instinct' categories were given. Reference to second level detail was the most frequently stated category and made up 47.09% of all decision rationale (excluding 'no response').

#### 4.4.2.2. Objective 4.2.2 A general comparison of Laboratory A and MPS Laboratory decision rationale

The proportionate composition of decision categories for Laboratory A and the MPS Laboratory is compared in Figure 4.10.

Decision factors reported for both groups fell into the same major categories of 'clarity', '1<sup>st</sup> level detail', '2<sup>nd</sup> level detail', 'area/type of mark'. The lesser represented categories of 'policy/situational factors', 'presentation may make a difference', 'comparative quality statement', and 'unsure' were also replicated over the two groups. One additional category was used with the MPS data.

Figure 4.10. demonstrates one key difference in the proportionality of decision categories between the two laboratories. Whilst the most represented decision category within each laboratory is '2<sup>nd</sup> level detail', there is a difference in the proportions of the second and third largest categories between the two laboratories. Laboratory A practitioners reported decision factors in relation to clarity of marks on more occasions than they did those related to 1<sup>st</sup> level detail (26% and 16% of decision outcomes, respectively). MPS laboratory practitioners, on the other hand, reported a higher proportion of decision outcomes in relation to 1<sup>st</sup> level detail (20%) than they did in relation to mark clarity (11%).

#### 4.4.2.3. Objective 4.2.3. Comparison of Decision rationale per crime context

Within Chapter 3 (3.4.3) it was established that there were differences in relation to fingerprint submission decision making according to crime category between the two laboratories. No statistically significant difference was determined in fingerprint submission decision outcomes according to crime context in relation to Laboratory A (see 3.4.3), whereas a statistically significant difference in decision outcomes was reported in relation to the MPS laboratory (Earwaker et al., 2015). Consequently, it was considered important to further explore the current data set to assess for any differences in decision rationale according to crime context within each laboratory. To do so decision rationale data in relation to the 20 borderline marks utilised in Earwaker et al. (2015) and in Chapter 3 of this thesis were analysed. Decision rationale stated by each practitioner in relation to each of those 20 fingerprints according to both categories of crime were collated from the larger overall decision sample, and the proportion of decision categories within this sample (according to laboratory and crime context) are presented in Figure 4.11.



Figure 4.10 - A comparison of the make-up of decision categories between Laboratory A and MPS Laboratory decisions

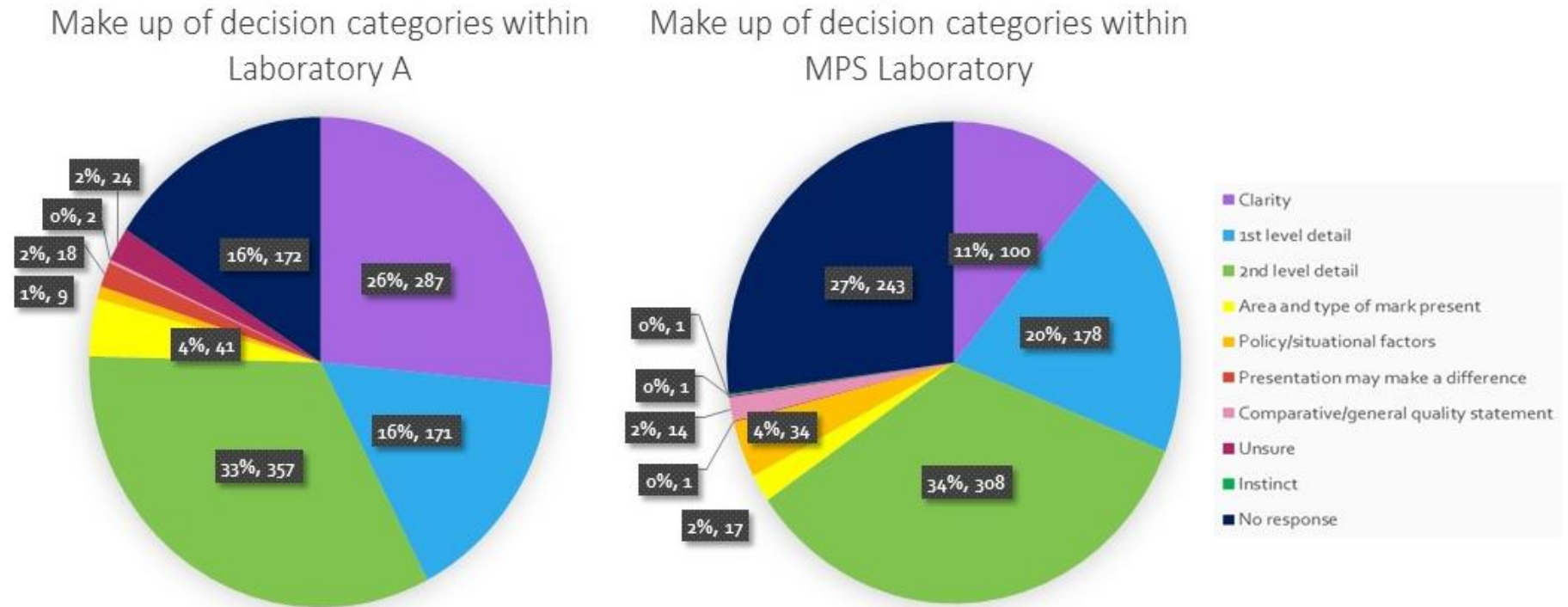


Figure 4.11 - A comparison of decision categories according to laboratory and crime context



Figure 4.11. illustrates consistency between laboratory and between crime contexts in relation to the type of major decision categories (other than no response); clarity, 1<sup>st</sup> level detail, and 2<sup>nd</sup> level detail. Rationale within the 2<sup>nd</sup> level detail category make-up the highest percentage of responses (other than no response) in each of the four cases, within proportions ranging from 31-37%. Between the two laboratories there can be seen to be a difference in the percentage value of the clarity and 1<sup>st</sup> level detail categories. Within the MPS laboratory, 1<sup>st</sup> level detail makes up a lower percentage of responses than clarity. Within Laboratory A, however, clarity responses account for a higher proportion than 1<sup>st</sup> level detail responses. The categories of decision rationale within Laboratory A are similar in relation to the two crime contexts, with a slight decrease in 'no response' (20%-12%), and a slight increase in the representation of the 'clarity category' (23%-30%) from the serious to the volume crime context. In the case of the MPS Laboratory there was a decrease in the proportion of the 2<sup>nd</sup> level detail and the clarity categories (31-37% and 15-10%, respectively) and an increase in the proportion of 'no response' responses from the serious to the volume crime category.

#### 4.4.3. Objective 4.3: Fingerprint examiner usability decision rationale

##### 4.4.3.1. Objective 4.3.1. Analysis of fingerprint examiner usability decision rationale

The decision rationale of 4 fingerprint examiners provided in relation to 30 fingerprint usability determinations were analysed. These usability determinations were made with access to good quality tenprint sets of the same source as the experimental images (as detailed in the Method, section 4.3), thus allowing the examiner the best possible chance to make an identification. During this process examiners were asked to make usability decisions through the process of identification and to provide their reasons for their decisions. Rationale were analysed through taking the same Ground Theory type approach as taken in relation to practitioner data (see section 4.3).

Grounded theory style analysis resulted in the identification of a total of 35 decision factors. These decision factors were grouped according to six decision categories. These were:

- Clarity/movement
- Pattern/1<sup>st</sup> level detail
- 2<sup>nd</sup> level detail
- Area of mark present
- Sequential ridge count and continuity or ridges
- Would get second opinion

The factors and frequency of factors making up each decision category are provided in Table 4.3. A description of decision factors for each decision category is provided below.

*Clarity/movement*

Decision factors within the ‘clarity/movement’ category cover aspects of the general clarity of the mark, such as ‘faint’, ‘dotty’, or ‘blurry’, and also factors which relate to movement or superimposition in the mark, for example ‘movement in mark’ or ‘superimposed’. Movement in mark is the most commonly reported rationale within this category. Clarity in ridges is also stated on a number of occasions. This category contained the largest number of decision factors (13).

*Pattern/1<sup>st</sup> level detail*

Only four factors were recorded within this decision category and there was a low representation of all of these factors. These involved the presence or absence of either a core or pattern within the fingerprint.

Table 4.3 - Decision factors and categories for examiner usability determinations

Clarity/movement	Freq.	Pattern/1 <sup>st</sup> level detail	Freq.	2 <sup>nd</sup> level detail	Freq.
clarity of ridges	16	pattern present	6	insufficient detail	28
movement in mark	19	core present	2	numerical value of characteristics	26
faint mark	16	no pattern	1	sufficient detail present	25
superimposed	15	Total frequency of all factors	9	ridge count too low	4
dotty	2			points in agreement	4
distortion	9			not many points	4
down twice	2			not enough characteristics	1
stretched	1			detail	1
Clear mark	4			enough detail in core	1
smudged mark	1			little detail in core	1
poor quality	1			few points in one area	1
detail clear	1			Total frequency of all factors	96
blurry	2				
Total frequency of all factors	89				
Area of mark present	Freq.	Sequential ridge count and continuity of ridges	Freq.	Would get second opinion	Freq.
area of mark present	6	charac in coincident sequence	3	would get seond opinion	2
Total frequency of all factors	6	ridges broken	8	Total frequency of all factors	2
		unable to count ridges	1		
		unable to find sequence of charac.	2		
		able to count ridges	1		
		few ridges	1		
		Total frequency of all factors	16		

### *2<sup>nd</sup> level detail*

The second level detail category was the second largest represented in terms of number of decision factors (with 10 decision factors). Key decision factors within this category were 'sufficient detail' (mentioned on 25 occasions), 'insufficient detail' (mentioned on 28 occasions) or reference to a 'numerical value of characteristics (26 occasions). An interesting factor within this category is 'points in agreement' which is referring to a comparative approach to quality assessment (i.e. there are points in agreement between the experimental fingerprint and the set of tenprint images provided), and is a clear difference in approach to quality assessment to that of laboratory practitioners.

### *Area of mark present*

Only one decision factor was assigned to this category. Examiners stated that the 'area of the mark present' had affected their decision making on six occasions.

### *Sequential ridge count and continuity of ridges*

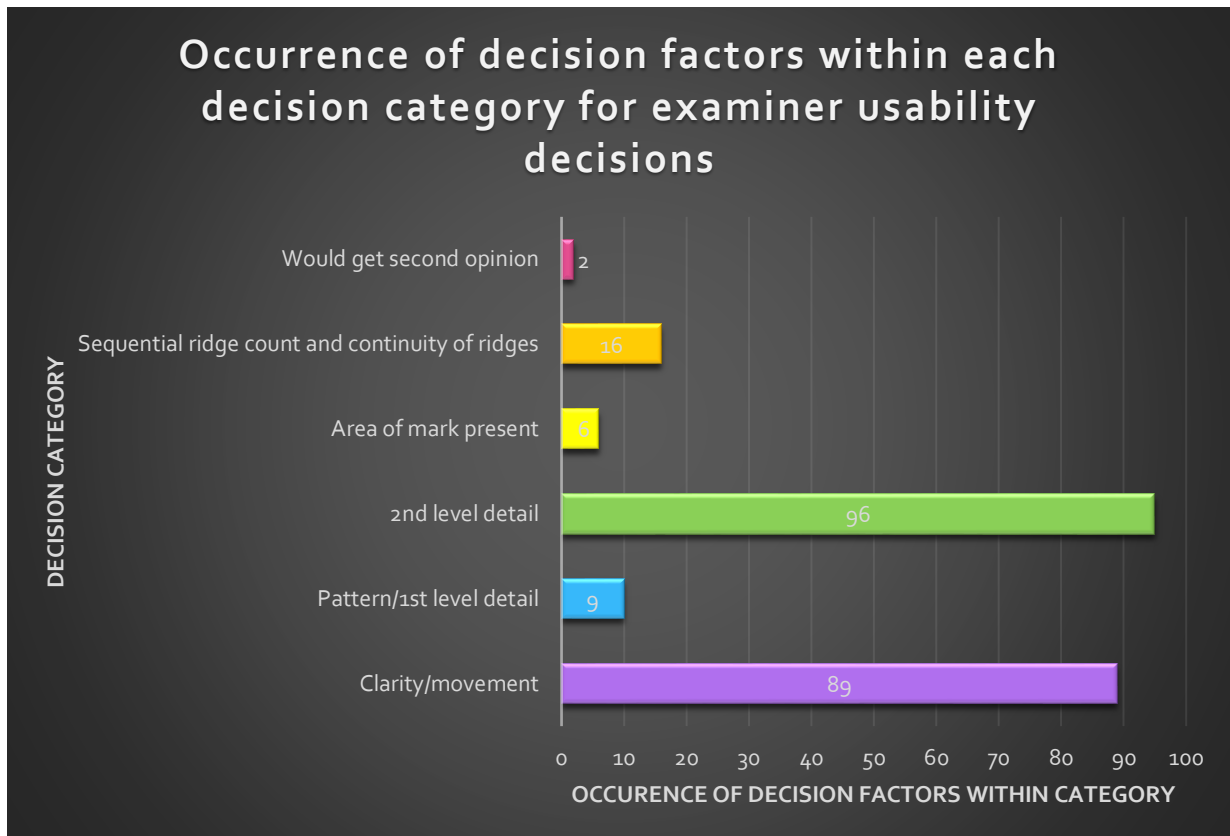
This category collates factors related to identifying the presence or absence of sequential ridges within a fingerprint and the sequential counting of characteristics. Six factors are included within this category. The inclusion of factors such as 'characteristics in coincident sequence' highlight the importance of continuity in the examiner decision process which was not so prominent within the practitioner quality assessment rationale.

### *Would get second opinion*

The decision factor 'would get second opinion' was given as a decision rationale on two occasions. This is similar to practitioner factors such as 'would get a peer's opinion' as included within the practitioner 'Unsure' decision category.

The total number of factors reported by examiners within each of the six decision categories are compared within Figure 4.12. This illustrates that '2<sup>nd</sup> level detail' was the most reported decision rationale by the examiners (on 96 occasions), and that 'clarity/movement' was also frequently reported (on 89 occasions). 'Pattern and 1<sup>st</sup> level' detail factors were reported considerably less often (on 9 occasions).

Figure 4.12 - Occurrence of decision factors within each decision category for examiner usability decisions

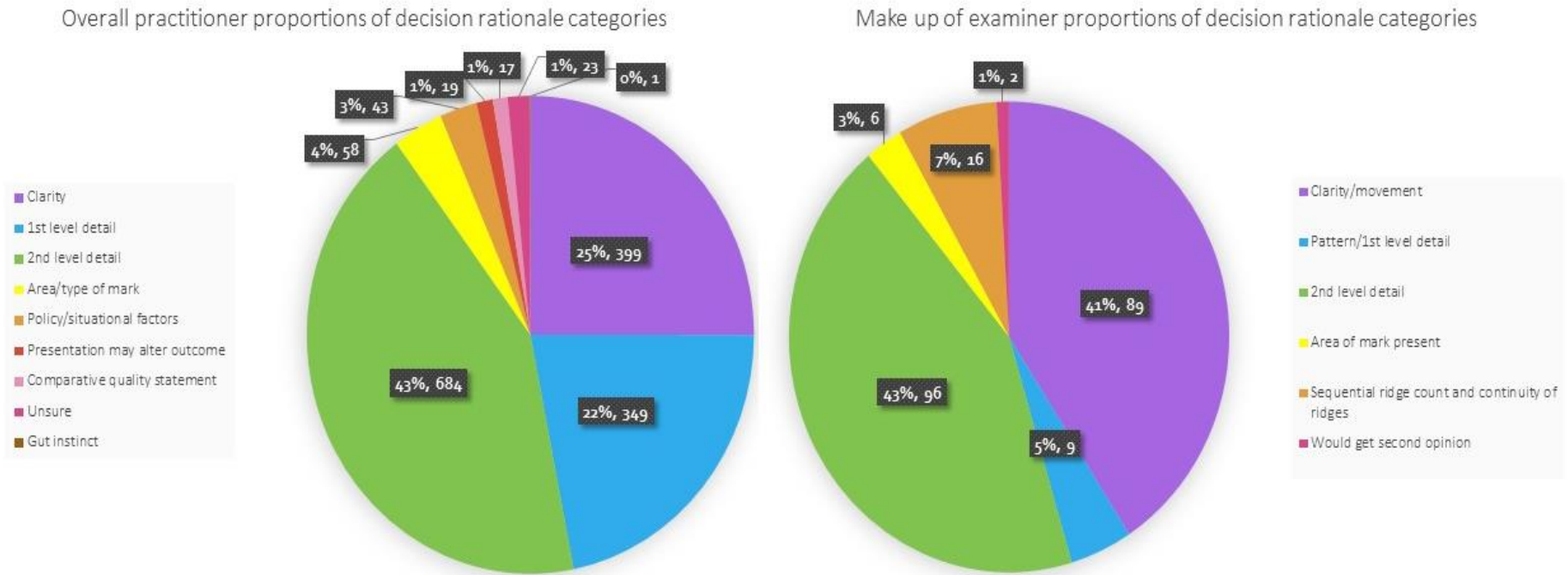


4.4.3.2. Objective 4.3.2. A comparison of decision rationale between practitioners and examiners

Practitioner decision rationale from both laboratories (Lab A and the MPS Laboratory) in relation to the same 30 experimental fingermarks examined with reference to examiner decision making in Objective 4.3.1 was collated. 'No response' factors were removed from this data set in order to allow a comparison of only decision related factors. The proportionate make-up of decision categories for the practitioner group and the examiner group are presented in Figure 4.13.

Due to the nature of the Ground Theory style analysis there are slight differences in decision categories between the two groups. However, the categories of 'clarity', '1<sup>st</sup> level detail', '2<sup>nd</sup> level detail' and 'area of mark' are present in both data sets and so can be compared. Both the practitioner and examiner groups can be seen to exhibit the same (and largest) percentage of rationale in relation to '2<sup>nd</sup> level detail' (43%). Within the practitioner group there is a similar proportion of 'clarity' (25%) and '1<sup>st</sup> level detail' (22%) decision categories. In the examiner group, however, 'clarity' makes up a large proportion of decision categories (41%), whilst '1<sup>st</sup> level detail' accounts for a small proportion (5%).

Figure 4.13 - A comparison of the make-up of decision rationale categories between practitioners and examiners



## 4.5. Summary of key findings

The key findings of empirical research within this chapter are summarised as follows:

Objective 4.1. To explore the rationale behind laboratory practitioner fingerprint submission decision making within Laboratory A.

Objective 4.1.1. A qualitative assessment of practitioner self-reported submission decision rationale within Laboratory A

- A broad range of rationale were reported by practitioners when deciding whether to submit or discard an experimental mark.
- Most commonly reported categories of rationale were associated with 2<sup>nd</sup> level detail and clarity of the mark.
- Within the category of '2<sup>nd</sup> level detail', factors which indicated that a count or threshold of minutiae was important made up a total of 78% of responses.

Objective 4.1.2. An assessment of inter-practitioner variation in fingerprint submission decision rationale with Laboratory A

The make-up of decision categories across individual practitioners was largely representative of the overall make-up of decision categories.

- There were some individual differences in the type of decision rationale reported.
- There were individual differences in the number of rationale reported.
- The 'no response' decision factor was more common amongst the responses of a minority of participants.

Objective 4.1.3. An assessment of practitioner fingerprint submission decision rationale according to experimental fingerprint

- There are similarities in the proportion of self-reported decision categories according to experimental mark.

Objective 4.2. An investigation of inter-laboratory consistency in practitioner decision rationale

- There were overall similarities in the decision categories observed across the two metropolitan UK police force laboratories.
- There was a difference in the relative proportion of the 'clarity' and '1st level detail' decision categories between the two laboratories, with the MPS laboratory reporting more decision



factors within the '1<sup>st</sup> level detail' category, than the 'clarity' category, and the reverse being true of Laboratory A.

- There were overall similarities in the make-up of decision categories given for both categories of crime by both laboratories.
- In the case of the MPS Laboratory there was a decrease in the proportion of the '2<sup>nd</sup> level detail' and the 'clarity' categories and an increase in the proportion of 'no response' responses from the serious to the volume crime category.

Objective 4.3. Exploring the relationship between practitioner submission decision rationale and fingerprint examiner usability rationale

Objective 4.3.1. An assessment of fingerprint examiner self-reported usability decision making rationale

- Fingerprint examiners reported a range of rationale for fingerprint usability decision making.
- Decision factors reported were most commonly related to the use of '2<sup>nd</sup> level detail', whilst 'clarity' and 'movement in the mark' was also a key category of rationale.

Objective 4.3.2. A comparison of fingerprint practitioner submission decision rationale and fingerprint examiner usability decision rationale

- There were overall similarities in the categories of decision rationale stated by examiners and practitioners.
- There was, however, a difference between the proportion of responses in relation to 1<sup>st</sup> level detail and the clarity of a fingerprint, with examiners relying more heavily on factors relating to mark clarity and practitioners relying more heavily on factors associated with the presence of pattern or 1<sup>st</sup> level detail and focusing upon continuity within a fingerprint.

## 4.6. Discussion

4.6.1. Objectives 4.1 and 4.2: Practitioner fingerprint submission making self-reported decision rationale

Consideration of low level responses

A broad range of rationale for fingerprint submission decisions (categorised as decision factors) were exhibited by practitioners within both Laboratory A and the MPS Laboratory, with a total of 93 different decision factors reported (16 unique to Lab A, 41 unique to MPS lab, and 36 reported within both laboratories). The Grounded Theory style analysis employed ensured that each of these decision factors was extracted from the data prior to grouping them at a higher level. This meant that the whole spectrum of reasons for mark submission decision making could be considered. Some of the lesser

reported decision factors may help to explore the thought processes surrounding mark submission for particular practitioners. The decision category 'Presentation could make a difference' contains three such interesting decision factors which shed light upon aspects of the mark submission process. 'Could be tweaked to get better contrast' is a decision factor reported by practitioners in both Laboratory A and the MPS Laboratory, this suggests an acknowledgment of the importance of imaging processes within mark submission. It could be considered that, as fingermark submission decision making occurs prior to fingermark image capture during laboratory workflow, the influence of the imaging process on the final fingermark to be submitted to the bureau is of key importance. Further research to empirically establish the effects of imaging processes on fingermark quality is therefore an important step to fully understanding the requirements of the practitioner submission decision. The decision factor 'better without a glass on' also alludes to the differences in the appearance of the mark according to the mechanism of viewing it. The factor 'expert may be able to see more' is also included within this category. This factor is interesting as it demonstrates that practitioners are considering differences in expertise between themselves and fingerprint examiners. This relationship in relation to the quantity of information within a fingermark will be further investigated in Chapter 5 of this thesis.

Practitioners within the MPS Laboratory refer to both 'unusual pattern' types and 'common pattern types' in their decision rationale, suggesting a preference for submitting rare pattern types for comparison. Whilst it is indeed the case that a rare pattern type may be easier to identify by an examiner, it is not the case necessarily that this identification is more valuable than an identification made in relation to a common pattern. It is also the case that there are differences in the commonality of fingerprint patterns across different populations (Kanchan & Chattopadhyay, 2006, Nanakorn et al., 2013, and Stamboulie et al., 2015), so basing a submission decision on the perceived commonality of a pattern type may not be an effective submission decision mechanism if used in isolation.

One decision factor provided during the experimental task was the use of 'gut instinct' during mark submission decision making. Whilst this may, perhaps, appear to be a non-scientific and worrying reason for determining whether or not to submit a fingermark, it may actually be, a very honest and helpful decision rationale and an accurate explanation of the decision process; using intuition to make a submission decision. Indeed, the use of intuition in decision making is not necessarily a negative concept. Dane et al. (2012) found that the effectiveness of intuition in analytical thought is increased with increased domain experience. The decision factor of 'gut instinct' could also suggest a lack of metacognitive awareness (knowledge of cognition) in the submission decision. Research has suggested that human expert decision makers often have limited metacognition and that some

cognitive activities are inaccessible to metacognition (Nisbett & Wilson, 1977). Metacognition of cognitive processes has been shown to be of benefit to decision making (Batha & Carroll, 2007). However, Hochberg (2014) linked increased metacognitive awareness to decreased diagnostic accuracy in physicians. A high occurrence of 'no response' when asked for decision rationale within both laboratories (16% and 27%) may also indicate difficulties in determining the cognitive processes involved. Further evidence for a possible lack of metacognitive ability within the practitioner group is, perhaps, apparent within the 2<sup>nd</sup> level detail decision category. Commonly reported decision factors within this category were 'sufficient' and 'insufficient' detail. Such a broad description of how practitioners had made their submission decision may suggest that they were unaware of the cognitive process that had led then to such a determination of sufficiency or insufficiency, or may, on the other hand, suggest the high importance of a cognitive threshold within the submission decision making process. Newell and Shanks (2014) highlight that metacognition information can be gleaned from participants, however, the open ended question asked to participants in the present study would not meet the sensitivity criteria detailed by Newell and Shanks (ibid). Therefore, there would be merit in extending this piece of qualitative research to further investigate metacognition around fingerprint submission in line with the method detailed by Newell and Shanks (ibid).

A high level consideration of decision categories

A higher level comparison of the make-up of decision categories between practitioners within Laboratory A and those within the MPS Laboratory demonstrates a similarity in the types of decision rationale reported. The most commonly occurring rationale within both laboratories is 2<sup>nd</sup> level detail (33% of decision factors within Lab A and 34% of decision factors within the MPS Lab), with Clarity (26%, and 11%) and 1<sup>st</sup> level detail (16% and 20%) also largely represented within both groups. The largest decision category as a result of rationale provided by Laboratory A (2<sup>nd</sup> level detail) was further investigated to examine in more detail the type of decision factors within this category. The decision factors were shown to belong to a number of sub categories relating to different aspects of 2<sup>nd</sup> level detail. 78% of factors within this category related to either a numerical value or count of minutiae (number/magnitude of minutiae, or ease of counting minutiae), or suggested a threshold of minutiae being reached (sufficient or insufficient minutiae). Such a predominant reliance on minutiae count in relation to a threshold to determine mark sufficiency for submission is an interesting finding. An example of a minutiae count decision rationale was '6+ points present so submit'. This statement clearly expresses the use of the two key components of Signal Detection Theory (Phillips et al., 2001) the detection of minutiae (quality indicators) within the marks, and the implementation of a threshold at which this number represents a mark of sufficient quality for submission. With so many

practitioners describing the use a numerical count of minutiae and the adoption of such a threshold it would seem that minutiae may play a key role in the decision process. Minutiae counts have, historically been of importance within UK fingerprinting. Until its abolishment in 2001 (Mackenzie, 2011) fingerprint examiners adopted a 16-point standard meaning that 16 points needed to be consistent between a suspect and exemplar fingerprint for a match to be said to occur. This historical reliance upon the presence of a high number of minutiae in a mark may go some way to explain an apparent reliance on a minutiae count as a mark submission mechanism by practitioners. Indeed, comparing the decision mechanisms of practitioners employed prior to the abolishment of the 16 point standard with those who began practicing after this change would be an interesting extension of this study. It may be that counting minutiae is a way in which the highly subjective task of mark submission (research undertaken within Chapter 3 highlighted that there is no formal procedure or training in relation to determining which marks to submit) can be made more objective. It may be that the objectifying of this decision through the use of a minutiae count is, indeed, an effective strategy for mark submission. Chapter 5 will examine this idea through further empirical study.

Data in relation to inter-practitioner differences in decision rationale demonstrates a representative proportion of the decision categories across the practitioner group. It is not the case that individual practitioners are adopting a consistent rationale for their decisions, for example one practitioner basing all of their decisions on 2<sup>nd</sup> level detail, whilst another basing all their decisions on mark clarity, rather that the decision rationale given by each practitioner has varied in different cases. This may have suggested that practitioners were using different mechanisms for decision making according to the experimental mark resulting in a mixed profile of rationale per participant. However, consideration of the mixed make up of decision factors for each experimental mark suggests that this is not the case. The range of decision rationale reported by each practitioner may be an indicator of the high level of subjectivity of the decision making process. It may be that the criteria for mark submission depends upon the most immediate quality indicator within the mark, for example if a pattern is clearly present then that may become the indicator used, instead of looking further at this pattern. On the other hand, it could be the case that the rationale provided do not actually reflect the true rationale for mark submission decision making, and that the submission decision is a subconscious process and that practitioners are actively seeking a rationale to report in the study. This would fit with views of the challenges of metacognitive processes in experts (Nisbett & Wilson, 1977).

An additional consideration is that practitioners may be using multiple cues when making their submission decision. Indeed, many practitioners stated multiple decision factors in relation to the same fingermark, often illustrating an evaluative process such as 'the clarity is good, but there are

insufficient characteristics present'. It could be that a sequential process of quality indicators is being utilised, for example, making an initial estimation of quality and if this is sufficient moving on to pattern, and then 2<sup>nd</sup> level detail, and so on. Support for the use of quality indicators (or attributes) in this way may come from the Decision by Sampling literature which would suggest the comparison of the quality rating of a series of quality attributes against those in working memory (Stewart et al., 2006) (the present experimental design should have ruled out comparison against attributes in the decision context as marks were presented in isolation). The order in which the present marks were presented could, however, have had an effect on the quality attributes held in working memory (in relation to the previously view mark), so the effect of the order of marks on the rationale provided may well be an interesting extension of the present analysis.

It is important to consider that the findings presented with this chapter relate to the decision rationale of practitioners working within two metropolitan UK fingerprint enhancement laboratories. Whilst the data represent the rationale of all available practitioners within these laboratories, they do not represent a sample of practitioners from across the breadth of UK laboratories. Due to the varying nature of the structure and working practices of police scientific support departments across the UK and due of the varying nature of crimes typically investigated, it may be the case that differences in approach would be identified by taking a wider practitioner sample. As such, the data cannot be said to be indicative of the decision rationale of practitioners across the UK, rather provides a starting point for further understanding of the fingerprint submission decision making process.

#### 4.6.2. Objective 4.3: A comparison of practitioner submission decision rationale and examiner usability decision rationale

One of the major challenges to improving fingerprint practitioner decision making is the absence of an objective criteria for submitting fingerprints to the bureau, resulting in a seemingly highly subjective decision. In order to improve the accuracy and efficiency of fingerprint submission (as shown to be problematic in relation to challenging fingerprints in Earwaker et al. (2015) and Chapter 3 of this thesis) it is essential to better understand how examiners are making their fingerprint usability determinations, and how this relates to practitioner submission decision mechanisms, as, in an ideal world, the outcomes of these two processes would be the same upon consideration of the same fingerprint.

Consideration of the decision factors reported by examiners illustrates a number of interesting decision factors that were not reported, or were reported differently within the practitioner group. Examiners commonly referred to the presence or absence of continuous and sequential ridge flow, which was lesser reported by practitioners. Examiners demonstrated differences in terminology from

practitioners, using phrases such as 'movement in the mark' and 'down twice' as opposed to 'smudged mark' and 'overlapping mark'. This use of more formal terminology may be a result of the formal training in mark analysis provided to examiners but not practitioners (Lagden, 2014). Overall there were similarities in the categories of decision rationale reported by the practitioner and examiner groups, with the categories of 'clarity', '1<sup>st</sup> level detail', '2<sup>nd</sup> level detail' and 'area of mark' present in both data sets. Both the practitioner and examiner groups exhibit the same (and largest) percentage of rationale in relation to '2<sup>nd</sup> level detail' (43%). Within the practitioner group there is a similar proportion of 'clarity' (25%) and '1<sup>st</sup> level detail' (22%) decision categories. In the examiner group, however, 'clarity' makes up a large proportion of decision categories (41%), whilst '1<sup>st</sup> level' detail accounts for a small proportion (5%). This is interesting as it suggests that practitioners may put a disproportionately large focus upon the use of '1<sup>st</sup> level detail' when making submission decisions, whereas clarity and movement are more important factors to an examiner making a usability determination. This difference illustrates the importance of communication between a laboratory and their bureau so that practitioners can gain a better understanding on the requirements of their submission decision. It is, however, important to bear in mind the considerable difference in the size of the practitioner (n=24) and examiner (n=4) datasets within this study. Further examination of the role of 1<sup>st</sup> level detail and mark clarity amongst a larger sample of examiners may be beneficial to add further weight to these findings.

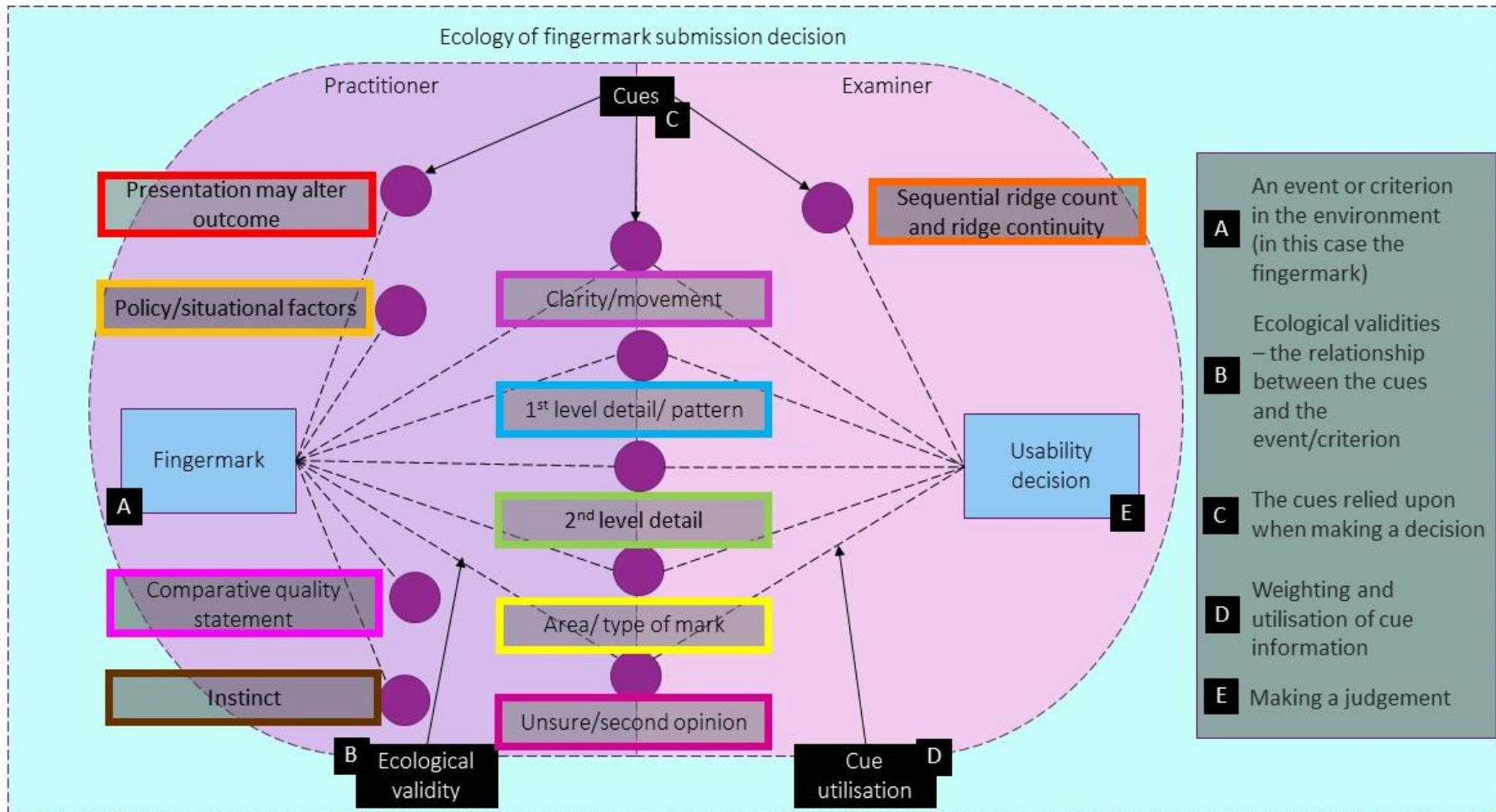
Previous research may suggest that the examiners preference for clarity and movement information may mean that this information is the most relevant in a fingerprint quality assessment task. Joseph and Patel (1990) found that expert physicians selected more relevant cues from a patient case history than novices. It would seem that examiners who are trained in quality assessment and whose role is more focussed upon the practical process of mark quality assessment prior to comparison would, indeed, be more able to select the more relevant quality indicators within a fingerprint. Similar findings have, indeed, been demonstrated within the crime and forensic science domains. Baber and Butler (2012) found that expert crime scene examiners focussed upon decision cues which were related to the ultimate use of recovered forensic evidence and Garcia- Retamero and Dhami (2009) found that burglars who were used to actually carrying out burglaries differed in their decision making to police officers who had knowledge of offender patterns but did not carry out the act of burglary.

As discussed in 4.1.1, the application of Brunswik's Lens model to the fingerprint laboratory fingerprint submission decision requires consideration of the factors (or cues) influencing the submission decision of the practitioner and those influencing the usability determination of the examiner. An extension of the novel application of Brunswik's Lens model is introduced in Figure 4.14.

Within Figure 4.14, the decision cues used by both practitioners and examiners are illustrated as connecting decision cues whereas those provided by only one group are illustrated as being detached from the complete decision process. Brunswik's Lens model also takes into account the ecology of the decision being made. In Brunswik's terms 'ecology' refers to the environment in which the decision is being made. Taking the ecology of decision into account is a departure from traditional psychological studies in which empirical studies tended to be taken out of their real world context, and findings tended only to be replicable in a laboratory environment. The present study sought to take into account the ecology of the mark submission process through an ecologically valid methodology. There were differences in the experimental task about which participants reported decision rationale between the practitioner and examiner groups in this study. Practitioners viewed fingermarks in isolation and were asked to make a submission decision, whereas examiners viewed marks alongside sets of tenprints of the same source and were asked to determine usability through an attempted comparison. As such, whilst the results cannot be said to originate from a direct comparison of performance in the same task, the data gained can be said to be highly ecologically valid, as these differences in task would be present within standard operational workflow. Resultantly this approach can be said to be more useful as it combines both the practical realities of the human expert within the process (the true ecology in which it exists) with the rigor of empirical study.

There are other key factors within the mark submission decision ecology that are important to consider, as these will be different for each mark submission decision. Such aspects of the ecology include the crime context, the presence or absence of suspects in the case, the time pressures associated with the case, the type of imaging used, and the financial requirements of the case. It is crucial that these aspects of the decision making are considered and that further research seeks to examine these factors and the influence that they have upon the mark submission process. The present study has successfully gained a naturalistic view of expertise within this decision making process, whilst working within an empirical approach. It is necessary to continue this approach to investigate the effect of all possible variables within the decision itself and within the ecology in this way, as forensic decisions such as this one do not exist in isolation or controlled laboratory conditions, and so it is important to be able to empirically assess the potential effect of all factors on the ultimate decision outcome.

Figure 4.14 - A novel adaptation of Brunswik's Lens Model including key decision cues gleaned from empirical study





#### 4.7. Conclusion

Performing a qualitative exploration and high level comparison of the cues used by practitioners and examiners when making the, essentially, equivalent decisions of mark submission and usability has led to a better understanding of the similarities and differences in the use of quality indicator cues between the two groups, and the relative importance of different decision cues within each group. These novel findings provide a unique insight into the metacognition of fingerprint laboratory practitioners and fingerprint examiners that can help to direct future research, and illustrates the value of successful collaboration between academia and operational forensic practitioners. The present study has highlighted that 2<sup>nd</sup> level detail is the most commonly stated rationale for fingerprint quality decision making within both the practitioner and examiner groups, and that a numerical count or threshold is a key factor in determining suitability for submission or comparison. Chapter 5 of this thesis will seek to further explore the relationship between the practitioner and examiner use of minutiae counts as a decision cue, through quantitative empirical study.

# Chapter 5 A comparison of examiner and practitioner minutiae counts, and an assessment of inter-practitioner variability

## 5.1. Introduction

### 5.1.1. The importance of minutiae counts within the mark submission process

Empirical study detailed within Chapter 4 of this thesis found that laboratory practitioners considered the quantity or count of minutiae present in a fingerprint to be an important factor in their fingerprint submission decisions. Practitioners frequently reported minutiae quantity or count as the rationale for a submission decision, even though it was established within Chapter 3 of this thesis that there was no documented procedure or criteria implemented for mark submission which would direct practitioners to focus upon minutiae count as a key submission criterion.

Given such an apparent reliance, in practice, upon the quantity of minutiae present within visualised ridge detail in order to make a submission decision it seems important to further investigate whether the use of such an objective criterion is, indeed, an effective mechanism for mark submission decision making. If this methodology is, indeed, an effective approach then there is a question as to whether this approach should be formally adopted in policy and as to whether the quantifiable nature of this approach could be exploited to allow the development of an empirically based, submission threshold of increased objectivity which could be readily adjusted to meet the requirements of the case or situation according to organisational requirements. For example, should there be found to be a quantifiable relationship between the minutiae counts achieved by examiners and practitioners then a numerical threshold for practitioner submission could be set which would ensure that the requirement (in terms of minutiae count) of examiners was met. The quantity of practitioner recognised minutiae required could be adjusted according to the varying requirements of the case or organisation (for example according to whether the latent mark was to be searched in an AFIS). Such a mechanism would, arguably, be a straightforward addition to laboratory practitioner processes as it would seem that a number of practitioners are already using such an approach but currently at varying thresholds, which may be an explanation for some of the inefficiencies in submission process identified in Chapter 3 of this thesis.

In order to investigate the potential of the use of such a decision mechanism there is first the need to empirically establish whether there is, indeed, any relationship between practitioner and examiner

minutiae count values, and, importantly, to establish whether there is a sufficient level of consistency in minutiae determination within practitioner and examiner groups.

#### 5.1.2. The use of minutiae counts by fingerprint examiners

Minutiae counts are recognised to be important in both the mark quality assessment and comparison processes carried out by fingerprint examiners (Ashbaugh, 1999). Indeed, minutiae count has been shown to be strongly associated with value determinations and the outcomes of comparisons made by examiners (Ulery et al., 2013, Ulery et al., 2014). This, combined with the knowledge of the use of minutiae count in mark submission decision making by practitioners, gained in Chapter 4 of this thesis, suggests that any relationship between the number and use of minutiae by fingerprint examiners and the number of minutiae determined by practitioners could, indeed, have the potential to be exploited to inform a more objective practitioner submission criteria.

A number of studies have examined the determination of minutiae by fingerprint examiners. Dror. et al. (2011) established that there was a lack of consistency in the 'Analysis' of marks (determination of minutiae) by examiners. These inconsistencies were both within examiners looking at the same mark on multiple occasions and also between different examiners looking at the same mark. The level of inconsistency was found to vary between fingermark and between examiner. Ulery et al. (2016) also found that minutiae count differed between examiners. Furthermore, they found that there was even greater variation between examiners in which minutiae were marked up during the 'Analysis' process, and, even where there was consistency in the number of minutiae marked up, there was still often inconsistency in which minutiae these were. Ulery et al (ibid) also established that the clarity of the minutiae in question had an effect on the level of reproducibility within the 'analysis' process (see 2.2.1) with a median reproducibility value of 82% in the areas of a fingermark in which the minutiae were clear, falling to 46% in areas in which the minutiae were ambiguous. Swoff et al. (2013) looked to establish a baseline of variation in minutiae detection by fingerprint examiners when dealing with the task in its simplest form. They provided examiners with only good quality, non-ambiguous marks and still found variations in the standard deviation of minutiae located. This demonstrates the existence of individual differences in this process, even in situations in which the identification of minutiae should be a straight forward and non-ambiguous task.

#### 5.1.3. Variation in performance in visual perception tasks

Variation in performance in similar visual perception tasks has also been found to occur within other domains. Potchen et al. (2000) assessed radiographers diagnoses of chest x-rays on a five-point diagnostic scale and found there to be considerable inter-observer variability in the scores achieved,

illustrating variation in the opinion of experts when using an objective scale. Bektas et al. (2009) also found intra and inter-observer variability in the use of the Fuhrman Grading System by pathologists to grade renal cell carcinoma. Within the five pathologists taking part in the study, consistency in results were achieved in only 48% of cases. This is an interesting finding when considering the minutiae determination task of practitioners and examiners, as the Fuhrman Grading System used is based upon the detection of the presence, and prominence of nucleoli. Similarity with a visual perception task within forensic science can be observed between the minutiae count task and the sperm scoring task carried out by forensic biologists. As part of the evidence analysis forensic biologists must count and record the number of visible sperm cells from vaginal swabs collected in cases of sexual assault. Tobe et al. (2015) investigated the reproducibility and subjectivity in the sperm scoring process finding sperm counting to be a highly subjective process with variation in the sperm count shown in relation to all of the experimental slides with no consensus between graders. It was also found that there was a higher level of agreement on sperm count the more extreme the number of sperm present, with less agreement on the 'average' sperm numbers. This suggests a possible contrast with the work carried out by Swoff et al (2013) who used good quality fingermarks as a benchmark for consistency in minutiae count. Whilst this is, indeed a valid benchmark as the marks used were clear and non-ambiguous, it may be (given the findings of Tobe et al. (ibid)) that greater consistency would be achieved using a smaller portion of a good quality mark in which a moderate quantity of minutiae were present as opposed to a whole good quality mark in which a large number of minutiae are likely to be present. The differences in variability of minutiae count according to the type of mark present is an interesting avenue of further study.

#### 5.1.4. The psychological basis for visual perception

Such a marked variation in the ability to determine the presence of apparently simple visual cues or signals may appear surprising. At a broad level such individual differences in visual perception can be explained by the constructivist theory of Gregory (1970).

Psychology has attempted to provide an explanation for the process by which the sensory inputs received by the sensory organs are converted into perceptions of that sensory input. In an attempt to do this views have been divided between two key, and fundamentally different, approaches that differ in the extent to which human perception relies directly upon the information present in the stimulus itself, versus the extent to which perception draws upon the expectations, experience, and prior knowledge of the perceiver. Gibson (1966) suggests an ecological, or direct, theory of perception which considers perception as a bottom up data driven process which occurs through the one directional transfer of information from the retina to the visual cortex. According to Gibson (ibid) this

process relies upon innate evolved mechanisms, and does not require any aspect of learning or hypothesising. Gregory (1970), on the other hand proposes a top down theory of visual processing. His theory suggests that contextual information and the experience of an individual aid in the understanding of a visual trace. Visual perception, according to Gregory (ibid) is a hypothesis based approach with the formation of incorrect hypotheses leading to errors in perception. It would seem that the constructionist theory of Gregory (1970) allows for the individual differences observed within minutiae count related processes within fingerprint comparison and within other domains in which visual perception is used as a diagnostic tool, due to individual differences in learning and experience.

#### 5.1.5. Potential differences between examiner and practitioner minutiae count

The high level of inter and intra variation in experts who carry out the perceptual tasks as part of their routine role in the fields discussed, may suggest there is similar variation in fingerprint practitioner in their minutiae detection task in which they could be considered to be experts, from the point of view of the regularity with which they carry out the task.

However, research within the fingerprinting domain may suggest a difference in the level of variation in minutiae count between laboratory practitioners and fingerprint examiners. Langenburg et al (2004) found a higher consistency between novices as opposed to fingerprint examiners in minutiae count in relation to 8 out of the 10 fingermarks in their assessment task. Schiffer and Champod (2007) also found that training and experience in the 'analysis' task decreased variability in the minutiae marking up (mark submission) process. It could be inferred from these findings that there may be a higher level of variation in practitioner minutiae count than in fingerprint examiner minutiae count, given that examiners will have received specialist training in the 'analysis' process, whereas such specific training is lacking for laboratory practitioners, and that a larger proportion of the working day of examiners will be taken up with the observation of minutiae, in contrast to the more varied role of the laboratory practitioner.

Both Langenburg et al. (2004) and Schiffer and Champod (2007) also found that the number of minutiae determined increased with training and experience. Suggesting, again that there could be a difference in the number of minutiae observed in the same mark by the practitioners who have less expertise in this specific area than examiners. This is in line with the anecdotal practitioner view expressed during the research outlined in Chapter 3 that when practitioners can see a certain number of minutiae (X minutiae) in a mark, examiners can see those and additional minutiae (X + Y).

#### 5.1.6. The difference between a numerical standard for submission and a numerical standard for identification

It is important when discussing comparative minutiae counts to be mindful of the connotations of numerical thresholds within fingerprinting and the distinction between a potential numerical standard for submission, and a numerical standard for fingerprint identification. The sixteen-point standard for fingerprint identifications was abolished in England and Wales in 2001 (Mackenzie et al., 2011), in part due to inconsistencies in examiner minutiae counts. The suggestion that there may be an advantage in the use of a numerical standard for the purposes of practitioner fingermark submission does not make the direct association that a numerical threshold of minutiae is necessary for an identification. Instead, the potential advantage of the use of minutiae as submission standard would be to allow submission according to a more (although not entirely) objective criteria. It is recognised that a major limitation of this approach is the lack of a numerical standard for identification within the UK and that, as such, the setting of any numerical threshold may itself lead to a false negative errors in cases in which other aspects of the mark could be exploited for the purposes of an investigation. This is particularly the case as there is an increased move within some schools of thought towards a more probabilistic approach to carrying out and reporting fingerprint identifications (Neumann et al, 2015). However, the operational reality would appear (from the results of the qualitative research discussed in Chapter 4 of this thesis) to be that practitioners are commonly already adopting such a numerical threshold in the process of making their submission decisions. As such, if there is found to be any correlation between practitioner and examiner minutiae counts, then this could be exploited to improve the performance, consistency, and transparency of the currently unofficial, but commonly adopted, procedure. However, a finding of a lack of correlation and a considerable variation in minutiae count within practitioners would suggest minutiae count to be an ill-advised mechanism for submission decision making and would suggest that further empirical study should be carried out to establish best practice methodologies for mark submission which practitioners should then be encouraged to adopt to reduce an inappropriate over reliance on minutiae count as a decision cue.

#### 5.2. Summary of objectives

The empirical study set out within this chapter will seek to examine the differences between examiner and practitioner minutiae count and to employ variability in minutiae count as a dependent variable. The objectives of the empirical study within this chapter are, therefore, as follows.

Objective 5.1 – an examination of inter-practitioner variability in fingermark minutiae count

To determine the level of variability in minutiae count between fingerprint laboratory practitioners, testing the hypothesis that there will be variation in practitioner minutiae count as found within fingerprint examiner minutiae counts (Ulery et al., 2016).

Objective 5.2 – a comparison of the variability of practitioners and examiners

To compare the level of variability in minutiae count between fingerprint laboratory practitioners and fingerprint examiners.

Objective 5.3 – a comparison of practitioner and examiner minutiae counts

To compare laboratory practitioner and examiner minutiae counts in relation to the same fingerprints within varying quality categories to test the hypothesis that examiners can see more minutiae in a fingerprint than laboratory practitioners.

Objective 5.4 – a comparison of variability according to fingerprint quality

To compare examiner and practitioner variability in minutiae count according to the quality of fingerprints presented.

Objective 5.5 – an assessment of inter-laboratory differences in minutiae count and inter-practitioner variability

To compare minutiae, count and inter-practitioner variability between two different UK fingerprint recovery laboratories.

### 5.3. Method

#### 5.3.1. Overview of method

This study required laboratory practitioner and fingerprint examiner participants to state the number of minutiae present within a series of images of ninhydrin developed fingerprints. The number of minutiae determined was then compared between the two groups, and variation within the two groups was identified and compared.

#### 5.3.2. Materials

##### 5.3.2.1. Production of experimental fingerprint images

A total of 60 fingerprint images were produced to form the experimental set of images for use in this study. 20 of these fingerprint images were taken from the experimental image set produced for use within experimental work outlined in Chapter 3. These 20 images were photographs of ninhydrin developed fingerprints which had been judged to be of 'borderline' quality by the researcher. For further detail relating to the production of these 20 mark images please refer to section 3.3. of this

thesis. An additional 40 marks were produced in order to allow the experimental set of images to be representative of the spectrum of possible qualities which may be observed in case work. In order to produce these additional marks three donors each deposited a series of latent marks on three clean sheets of white A4 Xerox Performer laser and inkjet printing paper (80g/m<sup>2</sup>). Marks were deposited with a variety of pressures and movements. Some marks were loaded with an amino acid substrate (Latent print reference pad: amino acid based™ (Lightening Powder: part no: 1-2791)) and depletion series of these loaded marks were made. Marks were deposited in a semi random manner in order to allow the development of a range of fingermarks that would be likely to include both clear, good quality marks, and poor quality marks lacking in detail. The inclusion of depletion series increased the possibility of achieving both good quality and poor quality marks.

All deposited marks were developed with Ninhydrin, as per CAST recommended procedure for the visualisation of fingermarks on a porous paper substrate (Bowman, 1998). Visualisation took place at the CAST chemical visualisation laboratory at Sandridge, Hertfordshire. Ninhydrin working solution was made up as per CAST recommended formulation (Bowman, *ibid*) and applied to each sheet of paper in a fume cabinet. Treated papers were placed in a Weiss Galenkamp Ninhydrin Oven at 80°C degrees and 65% relative humidity for 2 minutes.

Each sheet of treated paper was photographed in two sections using a Sony A77 camera on an Industria Fototechnica Firenze copy-stand with a 50 ml F2.8 macro lens. Papers were lit with Kaiser lights with daylight tubes. This ensured that all visualised marks were captured in a time efficient manner. The resultant images were stored electronically for subsequent processing and printing.

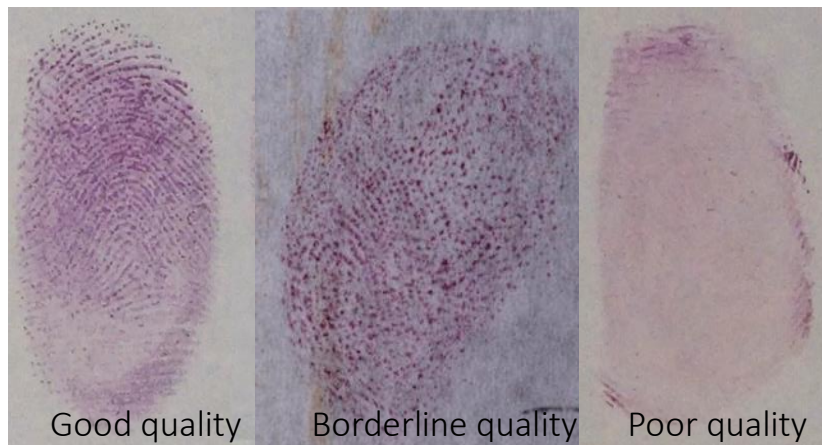
Images were viewed using Adobe Photoshop CS2. Twenty marks were selected by the researcher to form a 'good quality' image set. These marks were deemed to contain clear ridge flow and good contrast between the ridges and furrows present. An additional twenty marks were selected by the researcher to form a 'poor quality' image set. These marks contained broken up or poor contrast ridge flow and minutiae were not readily apparent. The selected mark images were cut in Photoshop from the original photographs and were pasted into one Photoshop file. The 20 'borderline fingerprint images (produced during previous experimental work) were also pasted into this file. The new images were rescaled to ensure that they matched the proportion of the original images. The quality of the scaled images was appraised to ensure they were still of the appropriate quality for inclusion in the image set. Examples of fingerprint images of each designated quality are provided in Figure 5.1.

Nine copies of the Photoshop file containing all 60 mark images were printed 1:1 within a UK police force photographic department on photographic paper, producing nine A4 printed photographs. Printed images were utilised in this study in order to ensure consistency in the image resolution



provided to participants. Had electronic images been provided to participants then these would have needed to have been viewed on monitors provided at the place of work of the participants. As police forces across the UK have differing facilities and equipment it is likely that participants would have viewed the images on screens of differing resolutions, limiting the reproducibility of the task in each location.

Figure 5.1 - Example of experimental fingermarks of each assigned quality ranking (enlarged image, not to scale)



#### 5.3.2.2. Production of experimental materials

Printed mark photograph sheets were divided into three pieces resulting in one sheet of 20 good quality marks, one sheet of 20 borderline quality marks, and one sheet of 20 poor quality marks from each of the nine photographs. Each set of 20 marks was numbered from 1-20 with all nine sets of marks of the same quality labelled in the same way. A random number generator ([www.random.org/sequences/](http://www.random.org/sequences/)) was used to produce three randomly generated sequences of numbers from 1 – 20. These sequences were used to create three different orders for the presentation of experimental marks. Each mark was cut out from the sheet providing small images containing only the mark of interest and was then placed in the appropriate randomly generated order. This resulted in sequences of experimental marks making up nine image sets as detailed in Table 5.1.

A further copy of each of these nine image sets was made (to allow one set of nine for each of the two scientific support departments to participate in the study) following the printing and labelling set out in 5.2.1.1.

Each set of images was collated in an A6 bound notepad with plain white pages with one image fixed with an adhesive 'dot' to each page in the random sequence generated. Presenting the images in this way ensured that only one image could be viewed simultaneously and increased the likelihood that participants would view the images in the order intended. A plain white background to each mark photograph ensured that there was no background interference to the marks and that each

participant viewed each mark within the same context. Using an adhesive ‘dot’ as opposed to an alternative adhesive ensured that the images were fixed securely and ensured that the adhesive did not cause damage to the appearance of the image over time.

Each fingerprint department was provided with a pack of experimental materials and instructions for carrying out the experiment, via a courier service.

Table 5.1 - Order of experimental image sets

Mark Quality	Mark Sequence Reference	Randomly Generated Order of Images	Image Set Reference
Poor	Sequence 1	13, 11, 18, 12, 19, 10, 14, 2, 9, 5, 4, 6, 8, 7, 3, 16, 17, 1, 15, 20	P1
Poor	Sequence 2	9, 5, 19, 18, 17, 15, 16, 20, 1, 8, 7, 14, 6, 2, 3, 4, 10, 13, 11, 12	P2
Poor	Sequence 3	6, 2, 7, 8, 12, 4, 1, 10, 20, 14, 9, 11, 19, 18, 17, 13, 16, 5, 15, 3	P3
Borderline	Sequence 1	13, 11, 18, 12, 19, 10, 14, 2, 9, 5, 4, 6, 8, 7, 3, 16, 17, 1, 15, 20	B1
Borderline	Sequence 2	9, 5, 19, 18, 17, 15, 16, 20, 1, 8, 7, 14, 6, 2, 3, 4, 10, 13, 11, 12	B2
Borderline	Sequence 3	6, 2, 7, 8, 12, 4, 1, 10, 20, 14, 9, 11, 19, 18, 17, 13, 16, 5, 15, 3	B3
Good	Sequence 1	13, 11, 18, 12, 19, 10, 14, 2, 9, 5, 4, 6, 8, 7, 3, 16, 17, 1, 15, 20	G1
Good	Sequence 2	9, 5, 19, 18, 17, 15, 16, 20, 1, 8, 7, 14, 6, 2, 3, 4, 10, 13, 11, 12	G2
Good	Sequence 3	6, 2, 7, 8, 12, 4, 1, 10, 20, 14, 9, 11, 19, 18, 17, 13, 16, 5, 15, 3	G3

The pack of experimental materials provided included:

- General instruction sheet outlining the experiment, timings and who should participate (see Appendix D.1)
- An instruction sheet detailing how to distribute experimental packs to participants (Appendix D.2)

- A record sheet for recording the dates packs were handed to each participant (Appendix D.3)
- An A4 envelope for each participant containing three A5 envelopes (one relating to each experimental task)
  - A5 envelope 1 containing:
    - Participant instruction sheet (Appendix D.4)
    - Participant information sheet (Appendix D.5)
    - Informed consent form (Appendix D.6)
    - Demographic information form (Appendix D.7)
    - Response sheet (Appendix D.8)
  - A5 envelop 2 containing:
    - Participant instruction sheet
    - Response sheet
  - A5 envelope 3 containing:
    - Participant instruction sheet
    - Response sheet
- A set of 9 photograph books (P1, P2, P3, B1, B2, B3, G1, G2, G3)

### 5.3.3. Participants

Participants were practicing Fingermark Visualisation Practitioners from two large metropolitan UK Police Force Scientific Support Departments (Laboratory A and Laboratory B) and Fingerprint Examiners from one UK Police Force Scientific Support Department (Laboratory B). A total of 25 practitioners (13 Laboratory A, and 12 from Laboratory B), and 21 examiners initially took part in the study.

### 5.3.4. Procedure

#### 5.3.4.1. Distribution of experimental packs

Both participating organisations were requested to appoint a coordinator to assist in the distribution of experimental materials to participants and ensure that the necessary timescales of the project were upheld. It was necessary to give this task to a scientific support stakeholder as opposed to the researcher as this meant that the participating organisations had the flexibility to accommodate the research around their workflow without the researcher needing to be present. The experimental design allowed flexibility in the exact timings for participation meaning that not all practitioners needed to participate at the same time, to ensure that the demands of operational casework could be

met at all times. The research coordinator was a manager within the fingerprinting unit of both of the forces who had daily contact with participants but would not be personally participating in the study.

Research coordinators were instructed to:

- Distribute the appropriate task envelope to the appropriate participant at the appropriate time (following the instructions written on the record of mark distribution). The study was designed to be counter balanced according to a) the order of the fingerprints presented (with each participant receiving one of three randomly generated orders), and b) the order in which tasks relating to the quality of the fingerprints were presented. The order and sequence (indicated by the book reference to be given to each participant on each occasion) was provided for the coordinators' use on the record of mark distribution
- Collect completed responses from participants upon completion of the task
- Liaise with the researcher in relation to any operational or staffing issues which may cause a delay in completion of the task or have the potential to result in withdrawal from the project
- Act as a first point of contact for participants, directing any unanswerable queries regarding participation to the researcher

#### 5.3.4.2. Experimental tasks

Participation in the study required the completion of three minutiae counting tasks. Each task involved counting and writing down the number of minutiae present in a series of 20 fingerprint images bound within a photograph book with one image per page. The quality of the fingerprint images (as determined by the researcher) differed between each task. The set of images was either of 'good', 'borderline', or 'poor' quality.

Participants were asked to read the instruction and information sheets and (when completing their first experimental task) to complete and sign the informed consent form and demographic information sheet provided.

Participants were instructed to look at each mark photograph within the book of experimental images provided and state in the corresponding space on the response sheet provided, how many minutiae they could see in the mark. They were asked to look at each image in isolation and in the order in which it was provided. Participants were told to carry out this task independently, within their normal working environment, using any standard equipment they would use to carry out such a task during casework. Upon completion of each task participants were directed to return their completed response sheet and photograph book to the research coordinator.

A minimum of 48 hours was left between experimental tasks in order to avoid any context effects from having recently viewed the experimental marks of a different quality. Counter balancing of the order of quality of marks presented was also intended to prevent any such unwanted context or order effects in the data.

## 5.4. Results

### 5.4.1. Aims of analysis

Data was analysed in order to meet the following objectives of the research, as outlined in the introduction to this study in section 5.1.7:

- Objective 5.1 – an examination of inter-practitioner variability in minutiae count
- Objective 5.2 – a comparison of the variability of practitioners and examiners
- Objective 5.3 – a comparison of practitioner and examiner minutiae counts
- Objective 5.4 – a comparison of variability according to fingermark quality
- Objective 5.5 – an assessment of inter-laboratory differences in minutiae count and inter-practitioner variability

### 5.4.2. Data preparation

Data was collated from all practitioners and examiners from both participating Fingerprint Departments. A total of three practitioners (two from Laboratory A and one from Laboratory B), and eight examiners had failed to complete all three experimental tasks due to operational requirements. As such, data relating to these participants was removed from the sample. In addition, one practitioner from each laboratory had not adhered to the instructions of the exercises and, instead of stating the total number of minutiae they had observed in each case, had set a threshold up to which to count minutiae, stating if a mark contained 6+ or 8+ minutiae, respectively. Similarly, one examiner had used a threshold of 20 minutiae in their responses. All data relating to these participants was removed from the sample. This left a data set containing 20 practitioners (10 from each of the two laboratories) and 12 examiners. The data set is provided within Appendix D.9.

### 5.4.3. Objective 5.1: An examination of inter-practitioner variability in fingermark minutiae count

Data analysis under this objective tested the hypothesis that there would be inter-practitioner variability in minutiae count, given the findings of variability in examiner minutiae count (Ulery et al., 2016).

Inter-practitioner variability according to fingerprint

Data gleaned from the practitioner minutiae count tasks was collected and collated according to each of the 60 experimental fingerprints provided to participants. Descriptive statistics related to the spread of the data for each fingerprint were calculated, and are presented in Table 5.2.

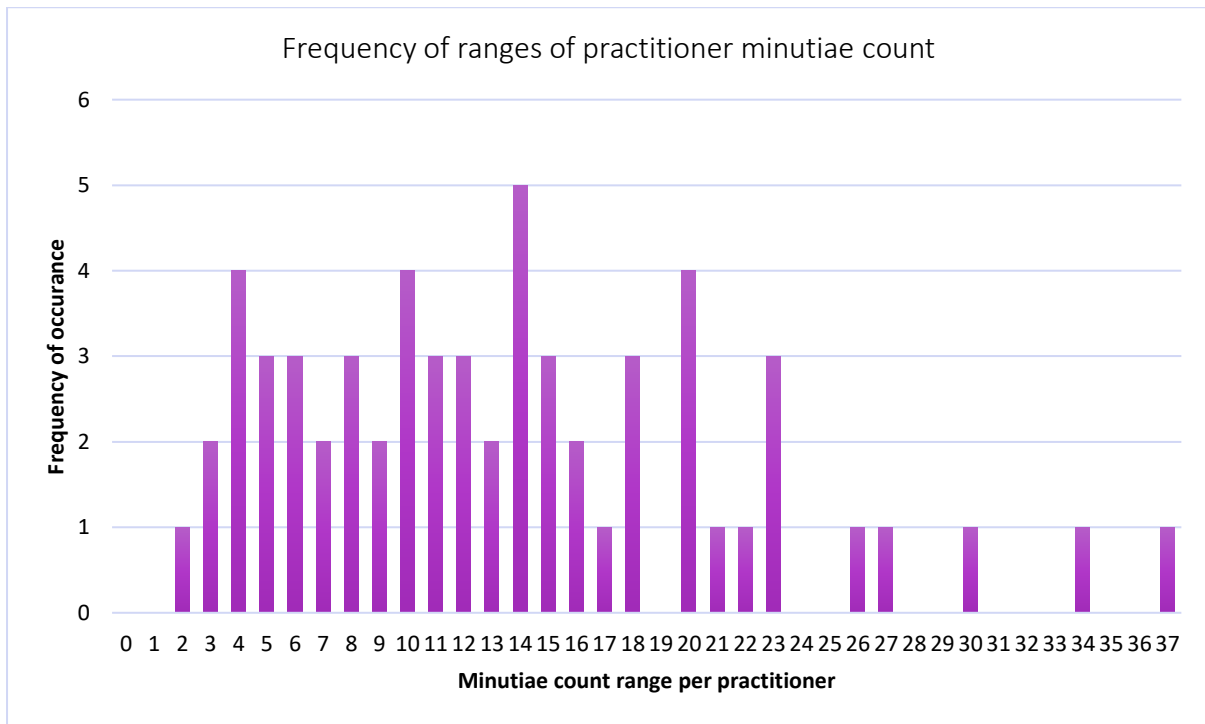
Table 5.2. provides the mean minutiae count, standard deviation, minimum, maximum, and range of minutiae counts for each mark. The data for each mark are grouped according to the quality classification the mark was given (either poor quality (left hand column), borderline quality (centre column), or good quality (right hand column)).

The data shows a considerable range in minutiae count in relation to many of the experimental fingerprints. For the majority of fingerprints at least one practitioner recorded an absence of any minutiae. In the case of the most extreme ranges minutiae counts of zero were recorded in relation to the same mark in relation to which minutiae counts as high as 22 were recorded. Ranges of minutiae count as high as 34 were recorded. A graphical representation of the spread of the range of minutiae counts in the data is provided in Figure 5.2.

Table 5.2 - Descriptive statistics for practitioner minutiae count

Poor quality marks						Borderline quality marks						Good quality marks					
Mark Ref	Mean Minutiae Count	SD	Min	Max	Range	Mark Ref	Mean Minutiae Count	SD	Min	Max	Range	Mark Ref	Mean Minutiae Count	SD	Min	Max	Range
P1	2.65	3.31	0	15	15	B1	2.85	2.32	0	7	7	G1	8.15	5.91	2	25	23
P2	0.7	1.03	0	4	4	B2	4.65	2.43	0	10	10	G2	17.1	7.93	5	42	37
P3	1.1	1.48	0	5	5	B3	6.8	3.91	2	16	14	G3	7.6	4.47	1	18	17
P4	0.9	1.02	0	3	3	B4	6.75	5.64	0	23	23	G4	15.55	8.25	4	30	26
P5	0.7	1.72	0	6	6	B5	7.35	4.08	2	16	14	G5	1.95	2.19	0	8	8
P6	1.15	1.76	0	6	6	B6	6.45	4.87	0	21	21	G6	6.1	4.08	1	15	14
P7	0.25	0.55	0	2	2	B7	2.1	1.62	0	7	7	G7	4.95	5.03	0	15	15
P8	0.4	0.82	0	3	3	B8	4.4	3.42	0	12	12	G8	4.35	3.17	0	11	11
P9	1.2	1.06	0	4	4	B9	7.15	4.58	0	20	20	G9	9.05	5.26	2	20	18
P10	0.55	1.19	0	5	5	B10	6.25	3.45	0	13	13	G10	7.75	5.99	1	24	23
P11	1.4	1.6	0	5	5	B11	5.8	5.44	0	22	22	G11	16.45	7.44	5	32	27
P12	2.8	3.33	0	12	12	B12	7.8	2.88	4	15	11	G12	6.15	4.28	0	16	16
P13	1.65	2.11	0	8	8	B13	3.2	1.82	1	7	6	G13	12.25	6.8	2	22	20
P14	2.35	2.48	0	10	10	B14	7.3	5.1	1	19	18	G14	2.7	3.53	0	14	14
P15	1.55	2.35	0	10	10	B15	4.6	3.42	0	15	15	G15	6.5	4.02	1	14	13
P16	0.35	0.93	0	4	4	B16	5.15	2.5	2	11	9	G16	2.9	2.67	0	10	10
P17	1	1.26	0	4	4	B17	7.65	7.35	1	31	30	G17	3.7	3.26	0	11	11
P18	2.65	3.47	0	14	14	B18	8.1	5.44	2	22	20	G18	6.6	3.97	1	17	16
P19	1.8	2.17	0	8	8	B19	5	3.03	0	12	12	G19	9.95	5.41	2	20	18
P20	2.2	2.59	0	9	9	B20	12.95	10.22	2	36	34	G20	9.25	4.97	2	22	20

Figure 5.2 - Frequency of ranges of practitioner minutiae count



The most commonly occurring range of minutiae counts was 14, occurring five times within the data set. Ranges of four, ten, and twenty were also common. The majority of ranges are clustered between two and 23, with the highest range of minutiae counts recorded being 37.

The standard deviations recorded in Figure 5.2 range from 0.55 to 10.22. The box plots presented in Figure 5.3 (A) provide a graphical representation of the spread of the data for the practitioner group. Marks are grouped according to their initial experimenter quality rating of poor, borderline, or good quality. Grouped in this manner, the data appear to be more spread in relation to fingerprints as mark quality (as per experimenter rating) increases. The effect of mark quality of minutiae count will be discussed further in relation to Objective 4.

#### Variability according to practitioner

Figure 5.4 illustrates the level of variation in the total number of minutiae observed in relation to all 60 of the experimental fingerprints by each of the 20 participating practitioners.



Figure 5.3 - Box plots of practitioner (Figure 5.3. A) and examiner (Figure 5.3. B) minutiae counts for all fingermarks

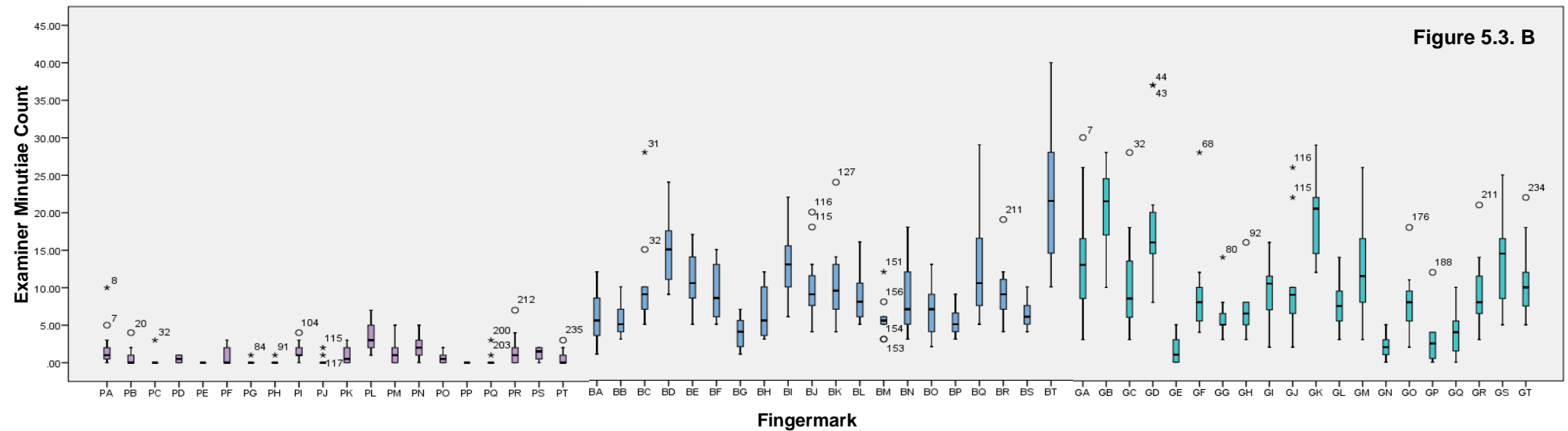
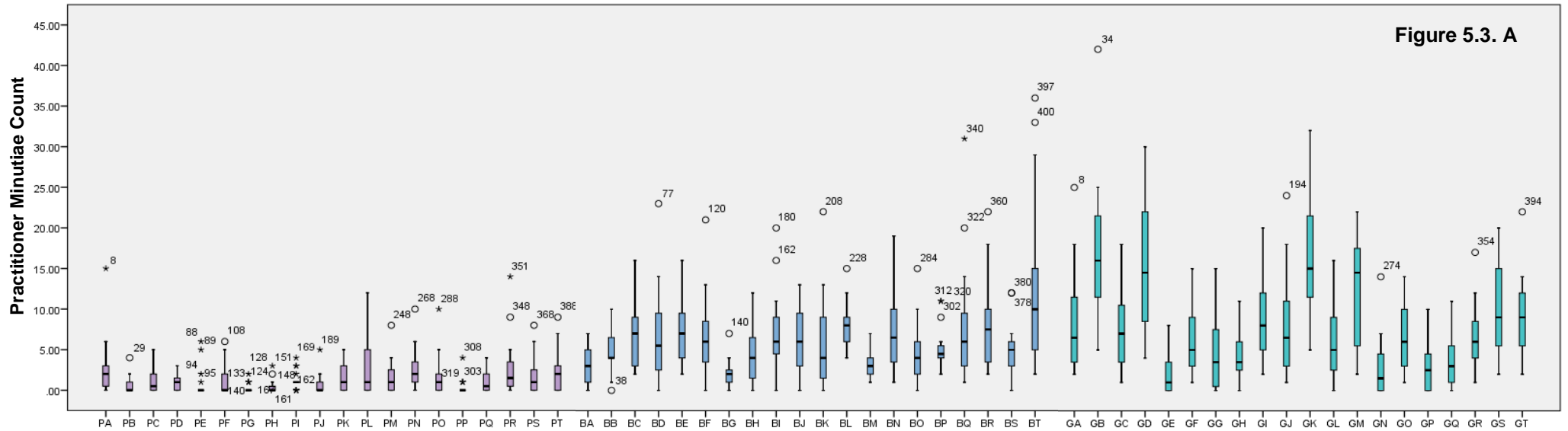


Figure 5.4 - Total number of minutiae observed by each participant in all experimental tasks

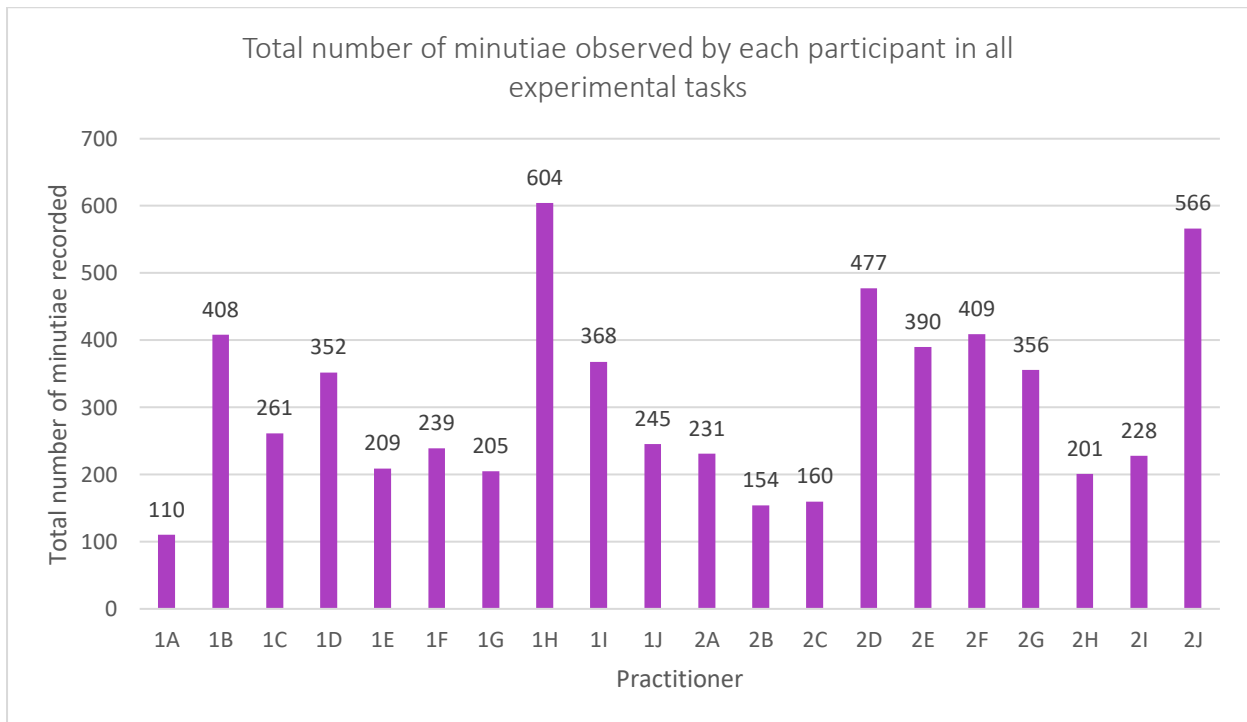


Figure 5.4 illustrates a high level of variation in total minutiae count between practitioners with a lowest total of 110 minutiae observed by one practitioner, compared to a highest total of 604 minutiae observed by another. 8 of the 20 practitioners reached a total of minutiae within the 200's.

Summary of Objective 5.1. results

Objective 5.1. set out to establish inter-variability in practitioner minutiae counts and to test the hypothesis that inter-variability in the minutiae counts of practitioners would, indeed be present, as inter-variability had previously been established within fingerprint examiners (Ulery et al., 2016). Inter-variation, was indeed found in relation to practitioner minutiae counts in the case of each experimental fingerprint.

5.4.4. Objective 5.2: A comparison of variability between practitioners and examiners

Data analysis under this objective compared the level of variability in minutiae count between fingerprint laboratory practitioners and fingerprint examiners.

Descriptive statistics were compiled in order to establish the extent of variation present within the fingerprint examiner minutiae count data. Mean minutiae count, standard deviations, and the range of minutiae counts given by examiners are presented in Table 5.3. Boxplots providing illustrations of practitioner and examiner variability according to each experimental fingerprint can be seen in Figure 5.3.

Table 5.3 - Descriptive statistics for examiner minutiae count

Poor quality marks						Borderline quality marks						Good quality marks					
Mark Ref	Mean Minutiae Count	SD	Min	Max	Range	Mark Ref	Mean Minutiae Count	SD	Min	Max	Range	Mark Ref	Mean Minutiae Count	SD	Min	Max	Range
P1	2	2.89	0	10	10	B1	5.75	3.25	1	12	11	G1	13.75	8.1	3	30	27
P2	0.75	1.22	0	4	4	B2	5.58	1.93	3	10	7	G2	20.5	5.3	10	28	18
P3	0.25	0.87	0	3	3	B3	10.25	6.21	5	28	23	G3	10.58	7.03	3	28	25
P4	0.5	0.52	0	1	1	B4	14.92	4.44	9	24	15	G4	18.92	9.25	8	37	29
P5	0	0	0	0	0	B5	10.83	3.81	5	17	12	G5	1.5	1.68	0	5	5
P6	0.92	1.31	0	3	3	B6	9.33	3.7	5	15	10	G6	9.3	6.39	4	28	24
P7	0.08	0.29	0	1	1	B7	3.92	1.93	1	7	6	G7	6	2.83	3	14	11
P8	0.08	0.29	0	1	1	B8	6.42	3.42	3	12	9	G8	6.92	3.29	3	16	13
P9	1.42	1.24	0	4	4	B9	12.83	4.61	6	22	16	G9	9.67	3.7	2	16	14
P10	0.25	0.62	0	2	2	B10	10.25	4.63	4	20	16	G10	10.5	6.78	2	26	24
P11	0.92	1.08	0	3	3	B11	10.58	5.3	4	24	20	G11	19.33	5.07	12	29	17
P12	3.5	1.88	1	7	6	B12	8.67	3.17	5	16	11	G12	7.67	3.31	3	14	11
P13	1.33	1.5	0	5	5	B13	5.83	2.37	3	12	9	G13	11.92	6.4	3	26	23
P14	2	1.54	0	5	5	B14	8.42	4.8	3	18	15	G14	2	1.48	0	5	5
P15	0.58	0.67	0	2	2	B15	6.75	3.17	2	13	11	G15	7.92	4.32	2	18	16
P16	0	0	0	0	0	B16	5.33	2.15	3	9	6	G16	2.92	3.26	0	12	12
P17	0.33	0.89	0	3	3	B17	12.83	7.7	5	29	24	G17	4.08	2.97	0	10	10
P18	1.5	2.11	0	7	7	B18	9.42	3.82	4	19	15	G18	9.67	4.68	3	21	18
P19	1.25	0.87	0	2	2	B19	6.33	1.87	4	10	6	G19	13.92	6.3	5	25	20
P20	0.58	1.08	0	3	3	B20	22.83	9.72	10	40	30	G20	10.75	4.97	5	22	17

Figure 5.5 - A comparison of practitioner and examiner standard deviation in minutiae count

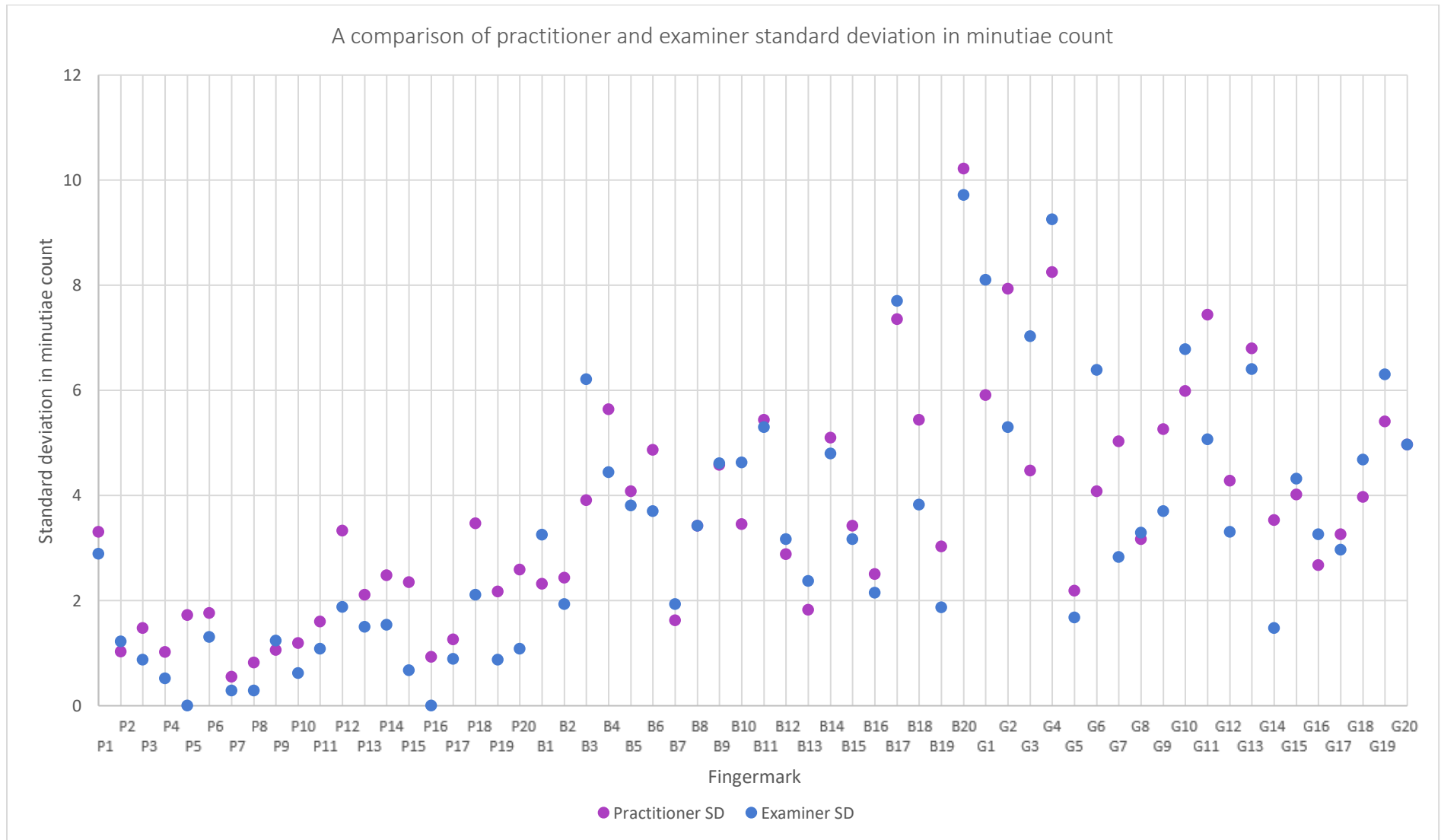


Figure 5.5 illustrates the standard deviations in minutiae count for practitioners and examiners. Standard deviation is used here as a measure of variation, with smaller standard deviations indicating higher central tendency or tendency towards the mean (less variation) and larger standard deviations indicating lower tendency towards the mean (higher variation). Standard deviation data is shown to be similar for the practitioner and examiner groups. The standard deviations in minutiae count per fingerprint can be seen to be larger within the practitioner population than in the examiner population in relation to 38 of the 60 experimental marks. Standard deviations are higher within the examiner group in relation to the remaining 22 fingerprints. Overall, however, there is a similarity in the mean standard deviation for the practitioner and examiner groups with mean standard deviations of 3.52 and 3.33, respectively.

Summary of Objective 5.2. results

Objective 5.2. set out to compare the level of variability in minutiae count between fingerprint laboratory practitioners and fingerprint examiners. Overall there was similar variability within the practitioner and examiner groups and on a per fingerprint basis practitioners demonstrated a larger standard deviation in minutiae count than examiners in relation to 38 of the 60 experimental fingerprints.

5.4.5. Objective 5.3: A comparison of practitioner and examiner minutiae counts

Having examined and compared the level of variation in minutiae count stated for each mark within the practitioner and examiner groups, minutiae count data was then analysed to establish how the actual minutiae counts of practitioners and examiners in relation to the same marks compared in order to test the hypothesis that examiners can see more minutiae in a fingerprint than laboratory practitioners due to a focus of training in the fingerprint 'analysis' process.

Figure 5.6 (provided in 5.4.6.) illustrates the mean minutiae counts of the practitioner and examiner groups for each of the 60 experimental fingerprints. Overall examiners gave the highest mean minutiae count in relation to 40 fingerprints, and practitioners gave the highest mean in relation to the remaining 20 marks. A further break down of this data according to fingerprint quality is provided in relation to Objective 5.4 in section 5.4.6.

Data were analysed to test for the presence of a statistically significant difference in the minutiae count of the practitioners and examiners.

#### 5.4.5.1. Statistical assessment of minutiae counts for all fingermarks

T tests were used in order to compare the mean minutiae counts reached by practitioners and examiners in this experimental task. Mean minutiae counts were compared in two ways; firstly by the use of a T Test to compare mean minutiae count per participant for between the practitioner and examiner group, and, secondly, by using a T Test to compare mean minutiae count for each experimental fingermark between the two groups. Repeating this analysis across the two different conditions allows an extent of generalisation from the participant sample to the participant population and generalisation from the fingermark sample, however, it should be born in mind that the sample of participating practitioners was not drawn from the population of UK practitioners at random.

A comparison of mean values of minutiae count per fingermark

Table 5.4 provides descriptive statistics for the practitioner and examiner groups when the data is analysed per experimental fingermark.

Table 5.4 - Mean minutiae count (per fingermark) values for practitioners and examiners

Participant type	N	Mean Minutiae Count	Std. Deviation in Minutiae Count	Std. Error Mean
Practitioner	60	5.1442	4.03153	0.52047
Examiner	60	6.7194	5.75009	0.74233

An independent-samples t-test was conducted to compare the mean minutiae counts per experimental fingermark for practitioners and examiners. There was no significant difference in the mean minutiae counts per mark for practitioners (M=5.14, SD=4.03), and examiners (M=6.72, SD=5.75);  $t(105.71) = 1.74$ ,  $p=0.085$  (two-tailed). The magnitude of the differences in the means (mean difference = 1.58, 95% CI: -3.37 to 0.22) was small (eta squared = 0.025).

A comparison of mean value of minutiae count per participant

Table 5.5 provides descriptive statistics for the practitioner and examiner groups when the data is analysed per experimental participant.

Table 5.5 - Mean minutiae count (per participant) values for practitioners and examiners

Participant type	N	Mean Minutiae Count	Std. Deviation in Minutiae Count	Std. Error Mean
Practitioner	20	5.1442	2.28691	0.51117
Examiner	12	6.7194	2.52422	0.72868

An independent-samples t-test was conducted to compare the mean minutiae counts per participant between the practitioner and examiner groups. There was no significant difference in the mean minutiae counts per participant for practitioners ( $M=5.14$ ,  $SD=2.29$ ), and examiners ( $M=6.72$ ,  $SD=2.52$ );  $t(30) = 1.82$ ,  $p=0.079$  (two-tailed). The magnitude of the differences in the means (mean difference = 1.58, 95% CI: -3.35 to 0.197) was moderate ( $\eta^2 = 0.099$ ).

Results show an overall lack of a statistical difference in mean minutiae count between the participating practitioners and examiners, with 9.9% of variance in minutiae count explainable by job role.

Summary of Objective 5.3. results

Objective 5.3. set out to test the hypothesis that examiners would determine the presence of a higher number of minutiae in fingermarks than practitioners, due to a higher level of training in relation to the 'analysis' process. Results demonstrate, however, a lack of significant difference in minutiae count between examiners and practitioners, refuting this hypothesis.

5.4.6. Objective 5.4: The effect of fingerprint quality on minutiae count and variability

Objective 5.4. set out to examine the effect of fingerprint quality on the numbers of minutiae counted and the variability of examiners and practitioners in this minutiae count.

Comparing mean minutiae count according to experimenter classification of mark quality

The mean minutiae counts of examiners and practitioners according to each of the three mark quality groups assigned by the experimenter are compared in Figure 5.6. In relation to the fingermarks assigned as 'poor' quality the mean minutiae count given by practitioners was higher than that given by examiners in 17 out of the 20 marks. In the case of the 'borderline' quality fingermarks, however, examiners provided the higher mean minutiae count in relation to all 20 marks. Examiners also provided the higher mean minutiae count in 17 out of the 20 'good' quality rated fingermarks.

The total values of minutiae observed by practitioners and examiners within each quality category of mark are illustrated in Figure 5.7.

Figure 5.6 - A comparison of practitioner and examiner mean minutiae count per fingermark

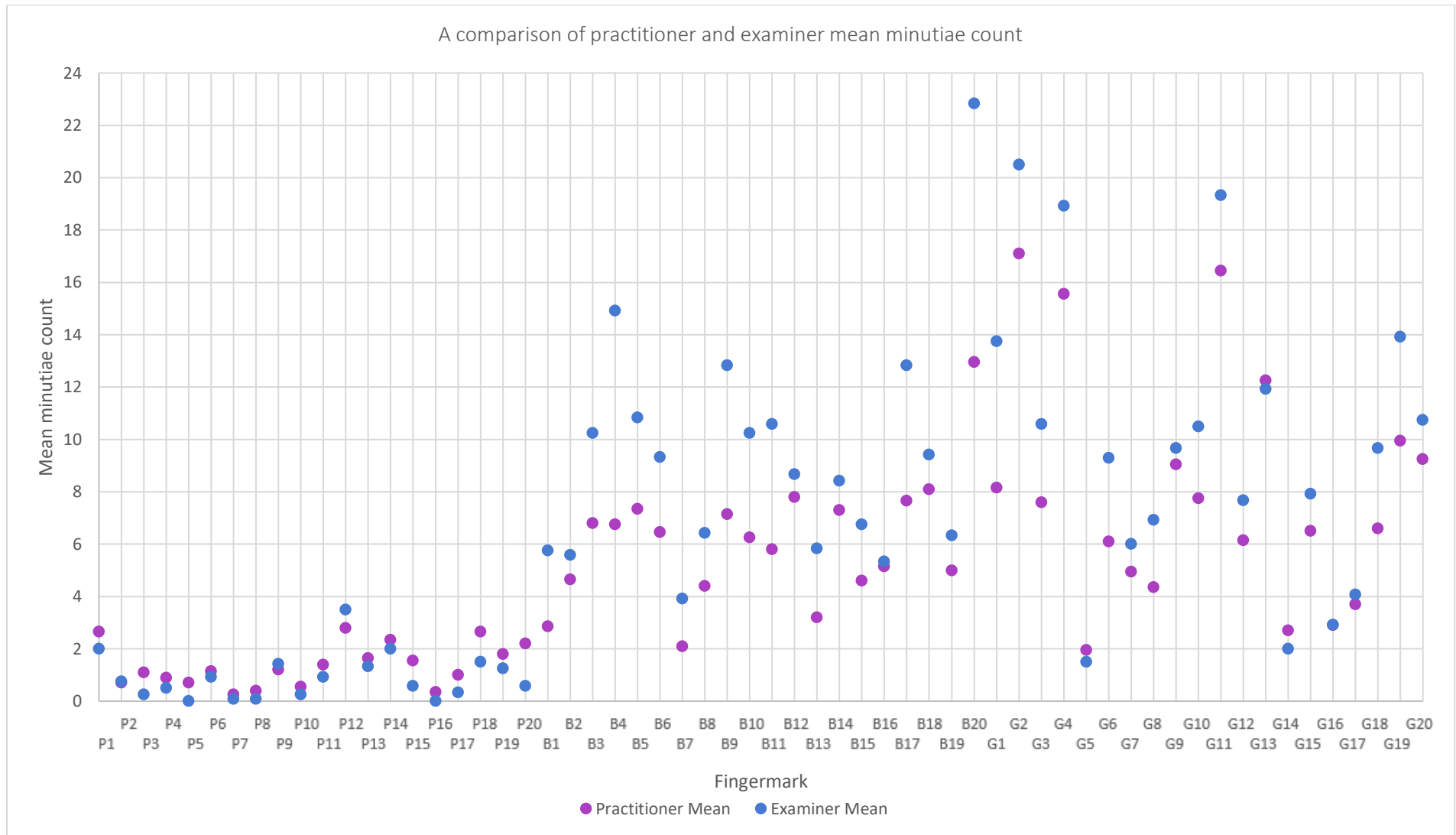




Figure 5.7 - Mean total of minutiae according to mark quality and participant job role

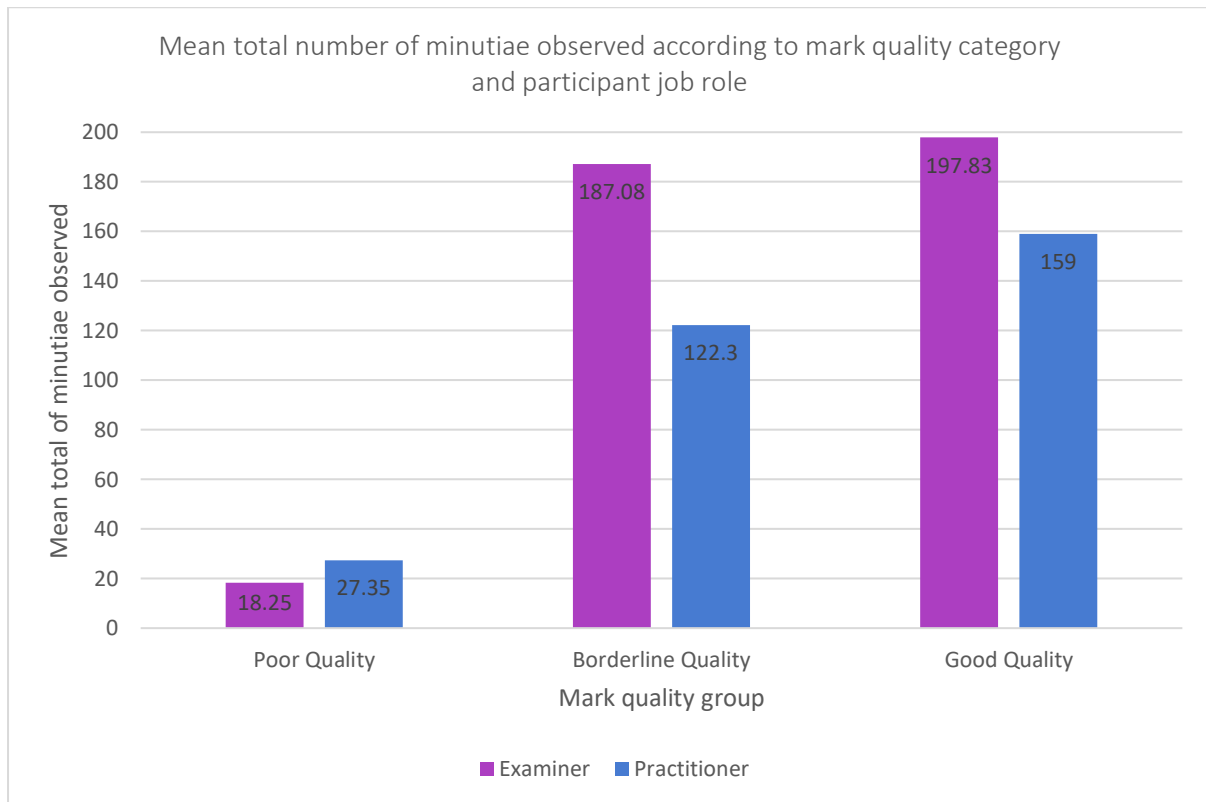


Figure 5.7 illustrates an increase in total minutiae observed as the researcher classification of quality increases within both the practitioner and examiner groups. Examiners observed more minutiae than practitioners when considering marks classified as borderline and good quality but observed less than practitioners in relation to poor quality marks. The biggest difference in minutiae count between practitioners and examiners was found to occur in relation to the borderline fingerprint group.

Assessment of a statistically significant relationship between mark quality and minutiae count and interaction effects between job role and mark quality

A two-way between-groups analysis of variance was conducted to explore the impact of job role and mark quality on minutiae counts. A two way between groups analysis was used in order to test whether there was an interaction between job role and mark quality on minutiae count, with job role and mark quality independent variables, and minutiae count the dependant variable. Fingermarks were divided into three groups according to their perceived general quality (by the researcher) (good quality, borderline quality, and poor quality). Levene's test of equality was significant ( $p=0.001$ ) suggesting that variance of the dependent variable was not equal across groups, as such, a more stringent significance level of 0.01 was set for analysing results (Pallant, 2007). The interaction effect between job role and mark quality was not statistically significant at a significance level of 0.01,  $F(2, 90) = 3.003$ ,  $p = 0.055$ . There was a statistically significant main effect for mark quality,  $F(2, 90) = 60.1$ ,

$p < 0.01$ , the effect size was large (partial eta squared = .138). Post hoc comparisons using the Tukey HSD test indicated that the mean minutiae count for the poor quality mark group ( $M=1.2$ ,  $SD = 1.2$ ) was significantly different from that of the borderline quality mark group ( $M=7.33$ ,  $SD = 3.58$ ) and the good quality mark group ( $M=8.68$ ,  $SD=3.83$ ). There was no significant difference ( $\text{sig } 0.01$ ) between minutiae counts for the good and borderline quality mark groups. The main effect for job role (examiner or practitioner),  $F(1, 90)=6.37$ ,  $p=0.013$ , did not reach statistical significance.

Using participant minutiae count as a quality benchmark for ranking the data set

The range, standard deviation and minimum and maximum number of minutiae determined by the practitioner group ( $n=20$ ) and examiner group ( $n=20$ ) were calculated and ranked according to the mean minutiae count of each mark of each of the participant groups. Ranking the data in this way allows consideration of the quantity of inter-variation in minutiae count according to the quality ranking of the marks in terms of minutiae count across the practitioner and examiner groups. In the absence of a 'ground truth' of minutiae count for this data, mean minutiae count is an appropriate method of ranking an overall practitioner and examiner view of quality in relation to minutiae. A similar approach was adopted in Tobe et al. (2015). Descriptive statistics for the practitioner group ranked according to practitioner mean minutiae count are provided in Table 5.6 and the same statistics in relation to the examiner group ranked according to examiner mean minutiae count are provided in Table 5.7.

Table 5.6 – Descriptive statistics for experimental marks ranked according to mean practitioner minutiae count

Mark Reference	Mean Minutiae Count	SD	Min	Max	Range
P7	0.25	0.55	0	2	2
P16	0.35	0.93	0	4	4
P8	0.4	0.82	0	3	3
P10	0.55	1.19	0	5	5
P2	0.7	1.03	0	4	4
P5	0.7	1.72	0	6	6
P4	0.9	1.02	0	3	3
P17	1	1.26	0	4	4
P3	1.1	1.48	0	5	5
P6	1.15	1.76	0	6	6
P9	1.2	1.06	0	4	4
P11	1.4	1.6	0	5	5
P15	1.55	2.35	0	10	10
P13	1.65	2.11	0	8	8
P19	1.8	2.17	0	8	8
G5	1.95	2.19	0	8	8
B7	2.1	1.62	0	7	7

Mark Reference	Mean Minutiae Count	SD	Min	Max	Range
P20	2.2	2.59	0	9	9
P14	2.35	2.48	0	10	10
P1	2.65	3.31	0	15	15
P18	2.65	3.47	0	14	14
G14	2.7	3.53	0	14	14
P12	2.8	3.33	0	12	12
B1	2.85	2.32	0	7	7
G16	2.9	2.67	0	10	10
B13	3.2	1.82	1	7	6
G17	3.7	3.26	0	11	11
G8	4.35	3.17	0	11	11
B8	4.4	3.42	0	12	12
B15	4.6	3.42	0	15	15
B2	4.65	2.43	0	10	10
G7	4.95	5.03	0	15	15
B19	5	3.03	0	12	12
B16	5.15	2.5	2	11	9
B11	5.8	5.44	0	22	22
G6	6.1	4.08	1	15	14
G12	6.15	4.28	0	16	16
B10	6.25	3.45	0	13	13
B6	6.45	4.87	0	21	21
G15	6.5	4.02	1	14	13
G18	6.6	3.97	1	17	16
B4	6.75	5.64	0	23	23
B3	6.8	3.91	2	16	14
B9	7.15	4.58	0	20	20
B14	7.3	5.1	1	19	18
B5	7.35	4.08	2	16	14
G3	7.6	4.47	1	18	17
B17	7.65	7.35	1	31	30
G10	7.75	5.99	1	24	23
B12	7.8	2.88	4	15	11
B18	8.1	5.44	2	22	20
G1	8.15	5.91	2	25	23
G9	9.05	5.26	2	20	18
G20	9.25	4.97	2	22	20
G19	9.95	5.41	2	20	18
G13	12.25	6.8	2	22	20
B20	12.95	10.22	2	36	34
G4	15.55	8.25	4	30	26
G11	16.45	7.44	5	32	27
G2	17.1	7.93	5	42	37

The tabulated data shows that mean minutiae count ranged from 0.25 to 17.1 minutiae across the sample of marks within the practitioner group. There was a trend for a higher standard deviation within the minutiae counts provided for each mark for marks in which the mean minutiae count was higher, demonstrating higher inter-variation when there was a higher number of minutiae available. The range of minutiae for each mark ranges from 2 to 37.

Table 5.7 - Descriptive statistics for experimental marks ranked according to mean examiner minutiae count

Mark Ref	Mean Minutiae Count	SD	Min	Max	Range
P5	0	0	0	0	0
P16	0	0	0	0	0
P7	0.08	0.29	0	1	1
P8	0.08	0.29	0	1	1
P3	0.25	0.87	0	3	3
P10	0.25	0.62	0	2	2
P17	0.33	0.89	0	3	3
P4	0.5	0.52	0	1	1
P15	0.58	0.67	0	2	2
P20	0.58	1.08	0	3	3
P2	0.75	1.22	0	4	4
P6	0.92	1.31	0	3	3
P11	0.92	1.08	0	3	3
P19	1.25	0.87	0	2	2
P13	1.33	1.5	0	5	5
P9	1.42	1.24	0	4	4
G5	1.5	1.68	0	5	5
P18	1.5	2.11	0	7	7
G14	2	1.48	0	5	5
P1	2	2.89	0	10	10
P14	2	1.54	0	5	5
G16	2.92	3.26	0	12	12
P12	3.5	1.88	1	7	6
B7	3.92	1.93	1	7	6
G17	4.08	2.97	0	10	10
B16	5.33	2.15	3	9	6
B2	5.58	1.93	3	10	7
B1	5.75	3.25	1	12	11
B13	5.83	2.37	3	12	9
G7	6	2.83	3	14	11
B19	6.33	1.87	4	10	6
B8	6.42	3.42	3	12	9
B15	6.75	3.17	2	13	11
G8	6.92	3.29	3	16	13
G12	7.67	3.31	3	14	11

Mark Ref	Mean Minutiae Count	SD	Min	Max	Range
G15	7.92	4.32	2	18	16
B14	8.42	4.8	3	18	15
B12	8.67	3.17	5	16	11
G6	9.3	6.39	4	28	24
B6	9.33	3.7	5	15	10
B18	9.42	3.82	4	19	15
G9	9.67	3.7	2	16	14
G18	9.67	4.68	3	21	18
B3	10.25	6.21	5	28	23
B10	10.25	4.63	4	20	16
G10	10.5	6.78	2	26	24
G3	10.58	7.03	3	28	25
B11	10.58	5.3	4	24	20
G20	10.75	4.97	5	22	17
B5	10.83	3.81	5	17	12
G13	11.92	6.4	3	26	23
B9	12.83	4.61	6	22	16
B17	12.83	7.7	5	29	24
G1	13.75	8.1	3	30	27
G19	13.92	6.3	5	25	20
B4	14.92	4.44	9	24	15
G4	18.92	9.25	8	37	29
G11	19.33	5.07	12	29	17
G2	20.5	5.3	10	28	18
B20	22.83	9.72	10	40	30

The mean minutiae count ranged from 0 to 22.83 within the examiner group. As per the practitioner group, there was a trend towards a higher standard deviation in minutiae count as the mean value of minutiae observed increased. The range of minutiae observed for each mark spans from 0 to 30.

Summary of Objective 5.4. results

Objective 5.4. set out to examine the effect of mark quality on minutiae count and variability in minutiae count. The number of minutiae observed by the examiner and practitioner groups was found to increase as the quality ranking of the marks increased. In addition, examiners were found to observe more minutiae than practitioners in relation to borderline and good quality fingermarks, whereas practitioners observed more minutiae in relation to poor quality marks.

5.4.7. Objective 5.5: Inter-laboratory consistency in practitioner variability and minutiae count

Minutiae counts stated in relation to the experimental fingermarks were compared according to which of the two participating laboratories practitioners were employed by, in order to test for differences in minutiae count between the two laboratories. Descriptive statistics for practitioner minutiae count, according to the laboratory in which they work is provided in Table 5.8.

Table 5.8 - Descriptive statistics for practitioner minutiae count, according to laboratory

Laboratory	N (marks)	Mean Minutiae Count	Std. Deviation in Minutiae Count	Std. Error Mean
Lab A	60	5.0017	3.88550	0.50162
Lab B	60	5.2867	4.35137	0.56176

A similar overall mean minutiae count can be seen between the two laboratories with laboratory B, demonstrating slightly higher variation in minutiae count (SD=4.35) than laboratory A (SD=3.89).

Statistical comparison of practitioner minutiae count according to laboratory

An independent-samples t-test was conducted to compare the mean minutiae counts per fingermark between practitioners employed by laboratory A and practitioners employed by laboratory B. There was no significant difference in the mean minutiae counts per participant for practitioners in lab 1 (M=5.00, SD=3.89), and lab 2 (M=5.29, SD=4.35);  $t(118) = 0.38, p=0.706$  (two-tailed). The magnitude of the differences in the means (mean difference = 0.29, CI = 1.78 to 1.21) was very small (eta squared = 0.001).

Examiners participating in this study were all employed with the same fingerprint bureau, preventing a similar comparison for the examiner dataset.

Summary of Objective 5.5. results

Objective 5.5. set out to compare minutiae count and variability between the two UK laboratories. Results demonstrate no significant difference in the number of minutiae counted between the two laboratories, but some difference in the level of variation in minutiae count.

## 5.5. Summary of key findings

The key findings of empirical research within this chapter are summarised as follows:

Objective 5.1: An examination of inter-practitioner variability in fingermark minutiae count

Key findings in relation to Objective 5.1 are as follows:

- There is considerable variation in practitioner minutiae count in relation to the same experimental fingermarks
- The range of minutia counts is often high with the most commonly occurring range of counts (in relation to five fingermarks) being 14
- There is considerable variation in total minutiae counts in relation to all marks stated by each practitioner. Total minutiae counts vary from 110 to 604 minutiae

Objective 5.2: A comparison of variability between practitioners and examiners

Key findings in relation to Objective 5.2 are as follows:

- Overall there was similar variability within the practitioner and examiner groups
- On a per fingermark basis practitioners demonstrated a larger standard deviation in minutiae count than examiners in relation to 38 of the 60 experimental fingermarks

Objective 5.3: A comparison of practitioner and examiner minutiae counts

Key findings in relation to Objective 5.3 are as follows:

- Overall there was no statistically significant difference between the minutiae counts of practitioners and examiners
- Mean minutiae count of examiners was, however, higher in relation to 40 of the fingermarks whilst higher mean minutiae counts were observed within the practitioner group in relation to the remaining 20 marks

Objective 5.4 – The effect of fingermark quality on minutiae count and variability

Key findings in relation to Objective 5.4 are as follows:

- The number of minutiae observed by both the examiner and practitioner groups increased as the quality ranking of the marks increased
- Examiners observed more minutiae than practitioners in relation to borderline and good quality fingermarks whilst practitioners observed more minutiae in relation to poor quality marks
- When per fingermark data was arranged according to mean minutiae count both practitioners and examiners illustrated greater variation as mark quality (in terms of minutiae count) increased

Objective 5.5 – Inter-laboratory consistency in practitioner variability and minutiae count

Key findings in relation to Objective 5.5 are as follows:

- There was no significant difference in the minutiae counts of practitioners according to the laboratory in which they worked
- Laboratory B exhibited a slightly higher variation in minutiae count than Laboratory A

## 5.6. Discussion

The present research investigated practitioner and examiner minutiae counts in relation to two key areas which may be important in the use of a minutiae count threshold as an increased objective methodology for practitioner mark submission; the level of variation within the two groups, and the presence of a relationship between the number of minutiae observed between the practitioner and examiner groups. The results of the empirical study presented (Objectives 5.1 – 5.5) are discussed in relation to those key research questions.

### 5.6.1. Variation in minutiae count

Results of this study demonstrated considerable variation in minutiae count within both the practitioner and examiner groups. The mean standard deviations in minutiae count were similar for the practitioners (3.52) and the examiner groups (3.33), with practitioners demonstrating a slightly higher overall standard deviation. In relation to individual fingermarks within the experimental set, the highest standard deviations found were 10.22 for the practitioner group and 9.72 for the examiner group (in relation to the same fingermark). There was a large range of minutiae counts provided, with the highest range being 37 within the practitioner group and the highest being 30 within the examiner group. The most common range of minutiae counts observed by practitioners was 14. Minimum minutiae counts within the practitioner group tended to be 0 - 3 minutiae for each mark, while some practitioners recorded minutiae counts in relation to the same marks in the 20s or early 30s. Minimum values of minutiae counts for examiners tended to be slightly higher (1-5) with similar maximum values.

These findings suggest that there is a similar level of variation between fingerprint laboratory practitioner minutiae counts as there is between those of fingerprint examiners. Whilst the standard deviations of both groups are not large in comparison to the mean values of minutiae count achieved, the ranges of minutiae count are substantial. The values of standard deviation and range of minutiae count are similar to those observed by Dror et al. (2011) within the group of fingerprint examiners who they assessed for inter-observer consistency. Inter-observer inconsistency amongst examiners were also observed by Evett and Williams (1996), Langenburg (2004), Langenburg (2009), Schiffer and Champod (2007), and Ulery (2016). Langenburg (2004) and Schiffer and Champod (2007) found that there was a greater variation in the minutiae counts of novices than of examiners. The present finding



of a similar level of variation in both the practitioner and examiner groups suggests that practitioners may be nearer, in terms of expertise in detecting minutiae, to examiners than complete novices. This would seem sensible as both examiners and practitioners are observing fingermarks and minutiae on a daily basis and the research within Chapter 4 of this thesis would suggest a strong consideration of minutiae by the majority of practitioners. Examiners and practitioner may, however, be expected to differ slightly in their ability to identify minutiae given that examiners are required to identify and compare the presence of minutiae to an exact extent during the ACE-V process, whereas practitioner seemingly only rely on minutiae as a guide to justify mark submission. As the practitioner minutiae identification task is not recorded there can be no feedback in the accuracy of the practitioner minutiae identification during routine casework (which would be beneficial in developing expertise (Ericsson & Lehmann, 1996), whereas the 'marking up' of examiners may well be 'verified' by another examiner providing such feedback during the course of casework. Equally, there are differences in the focus of training provided to practitioners and examiners which may have suggested a difference in level of expertise (Lagden, 2014). However, any such difference in expertise is not reflected in differences in variation in minutiae count between the two groups according to the findings of the present study. Further extension of this study to compare the location of the minutiae identified, such as that undertaken within examiners by Ulery et al. (2016) may well be valuable to assess differences in both the accuracy and the selection preferences of minutiae between practitioner and examiner groups.

#### 5.6.2. Relationships between examiner and practitioner minutiae counts

The present study illustrates that there was no statistically significant difference in relation to the overall number of minutiae observed by the practitioner and examiner groups (mean minutiae counts were 5.14 and 6.72 respectively). This, again, suggests that practitioners can be described as 'experts' as opposed to the novices examined by Schiffer and Champod (2007) who found that novices observed less minutiae than experts (fingerprint examiners in the case of their research).

#### 5.6.3. The effect of mark quality on variability and minutiae count

However, the present data does suggest differences in practitioner and examiner minutiae counts in relation to particular fingermarks and according to fingermark quality. In the present study the experimental fingermarks were divided according to their perceived quality by the researcher. 20 fingermarks were classified as 'poor quality' with broken up and unclear ridges, 20 fingermarks were classified as 'borderline' these marks were those about which agreement by fingerprint examiners in

terms of usability had not been reached (see Chapter 3), and 20 fingermarks were classified as 'good quality', with clear ridges and minutiae.

For both the practitioner and examiner groups the mean total number of minutiae observed increased as the quality rating of the fingermark group increased. Examiners observed more minutiae in the borderline and good quality fingermark groups than practitioners, but practitioners observed more minutiae in relation to the poor quality fingermarks. It could have been expected that examiners who could be considered to have a higher level of expertise in relation to the determination of minutiae would have excelled at the more challenging task of determining minutiae within a challenging mark. It would be interesting to carry out further analysis in this area to ascertain where practitioners and examiners have located minutiae in the case of such challenging marks. It may be that practitioners have lowered their threshold for detecting minutiae in these challenging cases and may, not, in fact have been correct in their determination of all minutiae counted, or may have erred on the side of caution when determining minutiae perhaps based upon the belief that when they can see X minutiae and examiner can see X plus Y minutiae, as disclosed in Chapter 4. Overall, there was found to be no significant effect of job role on minutiae count over the three quality ranks of marks. However, this finding was only marginally non-significant ( $p=0.013$ ) at the confidence level set of 0.01, suggesting (along with differences in effect according to the ranked quality of a mark), that slight differences in the make up of the quality of the mark set could lead to a significant difference in mean minutiae count between examiners and practitioners.

The present study found a trend towards higher standard deviations in minutiae count as the quality of fingermarks increased. This is demonstrated both according to the assigned quality ranking of fingermarks (as seen in Figure 5.3) and also when marks were ranked according to ascending mean minutiae count for both the practitioner and examiner groups (Table 5.6 and Table 5.7). Finding of increased inter observer variation in minutiae count according to mark quality (as rated in these two ways) is an interesting contrast to previous research which suggested that higher quality traces resulted in a lower level of variation in minutiae count and visual perception tasks. Within fingerprint 'analysis' Ulery et al. (2016) the clarity of minutiae present in a mark was found to affect the reproducibility of minutiae determination, with marks with clearer minutiae leading to a higher level of reproducibility. Swofford et al. (2013) also demonstrated a belief that less variation would occur when using good quality fingermarks by using only good quality, non-ambiguous fingermarks in an attempt to establish a baseline of variation in minutiae detection within fingerprint examiners. Within their sample of 'good quality' fingermarks, Swofford et al. (ibid) found that there was still variation in minutiae count, however this variation was larger in relation to marks in which breaks in ridges due to creasing were present. In another area of visual perception Tobe et al. (2015) found that there was

a lower level of variation in the sperm counting task undertaken by forensic biologists the higher the number of sperm present to count. The present study, in contrast to these findings suggested greater variation in minutiae count to be present in marks of better quality which had a higher number of minutiae present. However, it is important to bear in mind that both methodologies used to rank the quality of the marks within the present study were subjective themselves based upon either the judgement of the researcher, or the judgement of the participants (in terms of stated minutiae counts) due to the lack of an objective or computer based methodology of minutiae count which can outrank human performance in the case of all fingerprints (Neumann et al, 2016). As the capabilities of such technology improve it may be possible to extend the scope of this study and compare the actual and variation in minutiae count between examiners and practitioners in relation to prints of known 'ground truth' minutiae count. It would also be interesting to further examine the experimental fingerprints for the presence of creases causing breaks in ridges as it may be the case that, irrespective of overall quality determination, this causes a greater variation in minutiae count (as discussed by Swofford et al. (2013)). It may be that there was a higher level of creasing present in the good quality marks used in this study than in the poor or borderline quality groups leading to greater variation in minutiae count.

#### 5.6.4. Individual differences in visual perception

As suggested by Dror et al. (2011) a lack of consistency in minutiae count may "reflect the absence of objective and quantifiable measures as to what constitutes a minutiae" and may also be due to individual differences in visual perception and signal detection.

The idea of the use of top down processing (the idea that perception is a hypothesis based upon prior knowledge), as proposed by Gregory (1970) may help to explain some of the individual differences in minutiae count found to occur, as individuals will differ in their past beliefs, experiences, and expectations. Considerable research has been conducted within the field of medical imaging where accurate visual perception is crucial to reliable diagnoses. Krupinski (2011) carried out a review of key research within the field of medical image perception and highlighted two key aspects of image interpretation; visual perception, and cognition (or interpretation). Kundel et al. (1978) categorised false negative errors in medical imaging in three ways. Firstly as 'search errors' which occur when the trace of interest is not picked up within an image because it is outside of the useful visual field of the observer and so is not reported, secondly as "recognition errors' in which case the trace of interest is seen but is considered to be below the threshold of being ambiguous and so is not reported, and thirdly 'decision errors' where the observer does not consciously recognise the ambiguous features of the trace of interest or dismisses them and so the trace is not reported.

Research suggests that the importance of the search process (or the process of looking at the image) should not be overlooked. It was found that during visual perception tasks in both radiology (Kuprinski et al., 2006) and pathology (Mello-Thoms et al., 2011) observers gained an initial global impression of an image before focussing on areas of this image. Additionally, it was found that the amount of time spent looking at an image can influence success, with observers more likely to report a trait as being present when they have spent longer looking at the image (Kundel et al., 1989). Indeed, within the domain of forensic science the visual attention of crime scene examinations has begun to be investigated by considering the approach used by scene examiners (Baber & Butler, 2012, Butler, 2014), and the effect of visual attention on a global or local processing level has been considered within the field of facial recognition (Gao et al., 2011). This research was based upon psychological studies which have investigated the effect of priming observers to carry out visual perception tasks using either global or local processing (Navon 1977, Lewis et al. 2009, Forster 2008, Dale & Arnell, 2014). There is, indeed, scope for further work to apply such studies to the 'search' process by which practitioners and examiners visually approach a fingerprint, looking at the aspects of processing which may influence this approach and using technology such as eye tracking systems to establish individual differences in approach.

The field of Signal Detection Theory (SDT) can be applied to the idea of 'recognition errors' and 'decision errors' proposed by Kundel et al. (1978). SDT, similarly, divides the perception task in to discrimination ability and decision threshold, with discrimination ability the observer's ability to detect a signal from a background noise, and decision threshold the perceived line that turns a presumed negative (not a minutiae present) into a positive (minutiae present) (Phillips et al., 2001) (see Chapter 2 for a further explanation of Signal Detection Theory). SDT suggests that discriminating a trace from the background is more challenging in more ambiguous situations in which the signal to noise ratio is lower (such as in the case of a poor quality fingerprint). This being the case it could have been expected that the level of variation in minutiae detection would have increased, in the case of more challenging marks. However, it is important to bear in mind, that (especially in the case of the quality of marks being ranked according to mean participant minutiae count) there tended to be more minutiae present for possible detection in the better quality ranked marks, and it can be expected that there would be greater variation within this naturally larger sample. However, these traces were much less ambiguous and so given the task of 'Determine how many minutiae (points) are visible within the mark' it may have been expected that participants would be similar in their responses in these simple detection tasks. A possible similarity to this higher inter observer variation in a simpler signal detection task may come from research within the medical domain. Berbaum et al. (2010) report that when radiologists detect one abnormality on a radiograph subsequent issues are often missed. It may be that

practitioners and examiners, having detected many minutiae in the case of good quality, non-ambiguous marks feel, that the search for additional minutiae has become less important due to the number of minutiae recorded and so terminate their search, either as a conscious termination of the search, or due to subconscious faulty decision making. The point at which such a termination occurs may well depend upon an individual's own numerical threshold for submission, or past experiences, leading to greater individual differences in these less ambiguous cases in which more minutiae are available. The determination of a threshold of minutiae for the submission of a mark should be considered independently from the SDT terminology of threshold which in this application would be the point at which a signal is recognised as being a minutiae. An indication of a possible consideration of threshold for the submission of a mark is apparent upon consideration of the responses given by two practitioners and one examiner in the present study, which were removed from the final data sample. These participants had imposed their own cut off threshold on the minutiae counting task, one practitioner stating that this was the point at which they would terminate their count and submit the mark. The cut off values stated by the two practitioners were 6 and 8 minutiae, whilst the examiner had terminated their count at 20 minutiae. It may, therefore, be the case that personal submission thresholds played a part in the termination of the minutiae count process throughout the data sample, and that this led to increased variation in minutiae count in relation to the quality of marks in which this threshold was more often met.

Signal detection theory could be used to draw out more information in relation to the minutiae detection task. Relative operating characteristic analysis (ROC) (based upon SDT) can be used to further explore diagnostic accuracy and measure perceptual performance in such tasks (Mickes et al., 2012). Such an approach has been taken within human visual perception within medicine (Chesters, 1992, Berbaum et al., 1989, Somoza and Mossman, 1991, and Versteeg et al., 1998) and within the study of eyewitness testimony (Mickes et al., 2012). As such, the future application of this methodology to minutiae detection within practitioners and examiners may be valuable.

#### 5.6.5. Operational significance of these findings

The findings of Chapter 4 suggested that the quality of minutiae present in a fingermark was an important factor in the practitioner mark submission decision, and that, in practice, a numerical threshold may be being adopted by some practitioners. The purpose of the present study was to ascertain the potential value in the use of minutiae count as an objective methodology for practitioner mark submission, given the current apparent reliance on minutiae quantity in this process.

A significant finding from the present study in relation to the potential for the use of a minutiae count as a mechanism for mark submission, is the considerable variation between practitioners (and

between examiners) in minutiae count in relation to the same fingermark. The range of minutiae counts recorded for each mark is particularly important as it demonstrates the potential reality of adopting such a mechanism at an operational level. As the mark submission process requires the submission decision of one practitioner to correspond to the usability decision of one examiner, it is the individual minutiae counts and the potentially large discrepancy in these, that is important as opposed to an average across the two groups. The range of minutiae counts found to exist within the data could mean that if the lowest scoring practitioner was submitting fingermark B20 to the highest scoring examiner the practitioner would have recorded 2 minutiae present, while the examiner would see 40 minutiae in the mark, whilst the highest scoring practitioner submitting to the same examiner would have recorded 36 minutiae in the same mark. This becomes problematic when attempting to set a numerical threshold for submission. If the threshold was set at five minutiae, for example, in this example practitioner 1 would have discarded a mark in which an examiner would have identified a considerable number of minutiae, whereas it may well be the case that such a low threshold would result in false positive submission from practitioner 2 in relation to other fingermarks. Such variability would suggest that mark submission mechanism may not be possible on a one size fits all basis, and that naturally occurring individual differences need to be taken into account.

In terms of a relationship between the minutiae counts of practitioners and examiners, on average there was no significant difference between the minutiae counts of the two groups. This goes against the suggestion by practitioners that when they can see X minutiae examiners can see X + Y, in terms of an overall rule. However individual differences within the groups make it impossible to quantify a relationship between the minutiae counts of the two groups that is reliable on an operational level.

In addition, it is important to bear in mind that literature within fingerprint examination has also discovered a considerable level of intra-variation within fingerprint examiners (Langenburg, 2009, Dror (2011), further compounding the problem. A valid extension of this study would be an assessment of intra-variability within the participating practitioners in order to be able to consider this from both an examiner and practitioner point of view.

Ultimately it would seem the use of a minutiae count threshold as a sole mechanism for practitioner mark submission may not be successful given the level of variation inherent in determining the presence of minutiae. However, further training and feedback in minutiae count tasks may enable a higher level of consistency to be achieved as may the adoption of guidelines and standards in relation to this process (as suggested by Swofford et al. 2013). Such training may be widely applicable to fingerprint laboratories as there were similarities in the range and level of minutiae count between the two laboratories examined in this study. The findings of Fieldhouse and Gwinnett (2016) suggest

that such training and the successful creation of such guidelines may, however, be difficult to achieve. Further empirical study is required to better understand the decision mechanisms in play and how these can be most successfully exploited to increase the efficiency and success rate of the submission decision.

## 5.7. Conclusion

This empirical study has built upon the practitioner submission decision rationale reported in Chapter 4 that were found to focus upon the idea of sufficiency and numerical counts of minutiae. The findings of the present study show a considerable variation in practitioner (and examiner) minutiae counts in relation to the same fingerprint, and a lack of an overall relationship between practitioner and examiner minutiae counts. These findings illustrate the challenges in adopting a practitioner minutiae count threshold in the fingerprint submission process, particularly as there is the potential for considerable variation between the minutiae count of the individual practitioner making the submission decision and the corresponding individual examiner making the usability decision.

# Chapter 6 The effect of background mark quality on mark submission decisions

## 6.1. Introduction

### 6.1.1. Differences in mark quality assessment practices between practitioners and examiners

Research conducted and presented within Chapters 3 to 5 of this thesis has assessed practitioner and examiner decision making in relation to fingermarks presented in isolation. Whilst this is often representative of the mark context upon which a fingerprint examiner's quality assessment decision is based, it is not always representative of the context of a mark upon which a practitioner must make a submission decision (Forensic Science Special Interest Group, 2014). Routinely, laboratory practitioners make their mark submission decisions in relation to marks in situ upon exhibits which also often contain a number of other marks of a variety of qualities. This is a common approach to workflow as it allows image capture resources to be used to record marks that have been determined to be of good enough quality by the practitioner, as opposed to wasting resources capturing marks prior to the decision process.

The empirical study outlined within the present chapter examines the effect of background marks of varying quality on the submission decisions made in relation to 'borderline' quality (or ambiguous) fingermarks on an exhibit. The study aims to ascertain whether the quality of background marks may have the potential to lead to differences in the marking up process, and, ultimately, differences in the marks put forward for search or comparison against a person of interest in a case.

The difference between fingerprint submission decision making in situ upon an exhibit within the fingerprint laboratory versus the isolated mark usability decision made within the bureau may be a barrier to an efficient fingerprint submission process, if background mark quality does indeed effect mark quality assessment. If this were to be the case then this may provide some explanation for the inefficiencies in the process identified previously in this thesis (see Chapter 3). As such it is vital to identify, what, if any, effect on mark quality assessment the presence of background marks of varying quality may have, both from the point of view of the effect on the practitioner's mark submission decision, and view point of overall efficiencies in the mark submission process (i.e. the consistency of quality assessment between practitioners and examiners).



### 6.1.2. Contrast effects, assimilation effects, and the Inclusion/Exclusion Model

Within social psychology literature the effects of contextual information within evaluative judgements (such as determining the quality of a fingerprint) have been described as either contrast effects or assimilation effects, with the interaction between these effects framed within the integrative inclusion/exclusion model (Bless & Schwarz, 2010), as summarised in 2.8.6.1.

Contrast effects describe the presence of, for example, a good quality context stimulus causing a target stimulus to appear, in contrast, to be of lesser quality than they actually are. Assimilation effects, on the other hand, describe a situation in which a good quality context stimulus would make a target stimulus appear to be higher quality than in reality. Which of these two processes occurs is considered to be dependent upon a judges' belief on the relevance of the context information, information which is perceived as representative of the target, and norms which influence the perceived usefulness of the information (Bless & Schwarz, 2010).

Empirical study has provided an evidence base for the occurrence of assimilation and contrast effects in evaluative judgements, which may help in the generation of hypothesis in relation to the most likely effects to occur within in-situ fingerprint submission decision making. Judgements of the attractiveness of faces have been used to experimentally explore assimilation and contrast effects Geiselman et al. (1984). Judgement in facial features can be considered to share some characteristics with a fingerprint quality decision. Fundamentally both can be considered evaluative judgements, and both can be considered subjective decisions, involving elements of pattern recognition. Geiselman et al. (ibid) explored the effect of simultaneous presentation of faces on the attractiveness judgement of a target face. They found that assimilation effects occurred with the presentation of two attractive context faces leading to a higher attractiveness judgement in relation to the average attractiveness target face. They found this effect to persist when the target face was shown alongside only one context face, but that the target and context face had to be shown simultaneously for the assimilation effect to occur (Geiselman et al., 1984). This may suggest that the simultaneous presentation of good quality context fingerprints on an exhibit may lead to assimilation effects in the quality assessment of an ambiguous mark on that exhibit.

Weddell et al. (1987) compared the type of context effects shown when determining the perceived attractiveness of faces presented either in series or simultaneously. They found that when a series of images of faces were presented the attractiveness of those faces presented prior to a target face caused a contrast effect (i.e. when attractive faces were presented first in the sequence, the target face was considered to be less attractive). However, when faces were presented simultaneously (in pairs) an assimilation effect was shown to occur (i.e. when an attractive face was presented at the

same time as a target face the target face was considered to be more attractive) (Wedell et al., *ibid*). This difference in effect of sequential and simultaneous viewing of context information demonstrates the potential for differences in fingerprint quality assessment judgements made by consideration of a sequence of mark photographs (as per the typical task of a fingerprint examiner) in comparison to the quality assessment judgements made by consideration of a mark in situ surrounded by other marks (as per the typical task of a laboratory practitioner). This finding may also suggest that when individuals sample from working memory (when faces were presented in sequence) a contrast effect is more likely to occur, whereas when sampling from the decision context is used (simultaneous presentation) an assimilation effect is more likely.

Herr et al. (1983) examined the effect of priming on the determination of the perceived size of animals. They used a colour perception task to prime participants by showing them one of four different sizes of coloured area (extremely large, moderately large, moderately small, and extremely small). They then asked participants to judge the size of a real (unambiguous) or fictitious (ambiguous) animal. Assimilation effects were found to occur when moderate sized exemplars (primes) were given and the size of ambiguous animals were judged. Contrast effects were found to occur when extreme exemplars were provided and ambiguous animals were judged, and also when all unambiguous animals were judged, irrespective of the magnitude of exemplars (Herr et al, *ibid*).

This study suggests the potential for the magnitude of quality of context fingerprints to have an impact upon the type of context effect observed. Within forensic case work it is common to develop a range of quality marks present in a range of quantities across the surface of an exhibit as such it is of interest to investigate the impact of a range of proportions of good and poor quality context marks on the submission decisions in relation to target marks. This study may suggest that contrast effects would be shown to occur when a mark background is either of very high quality or of very low quality when dealing with an ambiguous target mark. Ambiguous target marks are of particular interest within fingerprint submission as it is in ambiguous situations that cause the most uncertainty within decision making (Kahneman and Miller, 1986), and thus are the type of marks that this thesis concentrates on. An interesting consideration based upon the findings of Herr et al. (1983) is the quantity of background information considered by the decision maker when making this comparative judgement. An exhibit, for example, consisting of 80% good quality (useable) fingerprints and 20% poor quality (not usable) fingerprints, could be considered to be an 'extreme' exemplar of an exhibit made up of good quality marks. Alternatively, it could be the proportionate quality of individual marks that could be considered to govern how extreme the context exemplar is considered to be. For example, context fingerprints with a more obviously higher quality rating (e.g. many more minutiae visible) may provide a more

extreme context exemplar causing a contrast effect, whereas those that are good quality but less obviously so may be examples of moderate exemplars and cause an assimilation effect.

Schwarz et al. (1991) investigated assimilation and contrast effects in part-whole question sequences. They found that a preceding question asked before a general question which was not seen to be part of the same conversation, led to the preceding question priming the answer to the general question and an assimilation effect occurring. When however, the preceding question was deemed to be linked to (be part of the same conversation as) the general question, then this preceding information was deemed not to be useable in establishing the answer to the general question as this information had already been disclosed in the same conversation and a contrast effect occurred (Schwarz et al, 1991). However, it was also found that a series of specific questions preceding the general question (for example being first asked specifically about marriage satisfaction and then generally about life satisfaction) caused the general question to be viewed as requiring a summary judgement based upon the preceding questions, causing an assimilation effect. This suggests the potential importance of connectivity in the decision sample and the norms in relation to considering a mark or considering an exhibit for influencing the type of context effect demonstrated within fingerprint submission decision making.

### 6.1.3. Decision by sampling

There are many examples of psychological theories based upon relative judgement, decision by sampling being an example of such a theory. Decision by sampling (DbS) literature (as outlined in 2.8.6.1) assumes that individuals make decisions through a series of comparisons between attribute values within an available sample (Stewart, 2009).

Stewart et al. (2006) suggest that there are three stages to this process. The example provided in Stewart (2009) illustrates the three stages of the decision by sampling process for making a choice between two options, (a) and (b): (a) being a 20% chance of £4,000 otherwise an 80% chance of £0 or (b) being a 25% chance of £3,000 otherwise a 75% chance of £0. Table 6.1 summarises each stage of the decision by sampling process for this decision (Stewart 2009), and, additionally, applies each stage of this approach to the fingerprint submission decision.

When considering decision by sampling in relation to mark submission, there can be seen to be two different prospects to the decision being made, either a) to submit the mark or b) to not submit the mark. The process of considering and selecting the target attribute (the attribute of the decision about which to make a comparison) is more challenging than in the financial chance option provided, as the attributes need to be defined and valued by the decision maker. The results of the empirical study

provided in Chapter 4 of this thesis give an insight into the relative importance of different target attributes (or quality indicators) within a fingerprint. These findings may suggest that a common attribute considered during decision by sampling would be the number of minutiae present in the target mark in comparison with the number of minutiae present in marks in the decision context or available in working memory. However, it could also be the case that a different attribute is used, a combination of different attributes is applied, or that it is a series of binary rankings of 'overall perceived quality' that occur until a threshold is reached.

An example of applied research into decision by sampling can be found within the medical and public policy literature. Wood et al. (2012) investigated perception of risk in relation to own alcohol intake through sampling of the alcohol intake of others and numbers of recommended units of alcohol consumption. The study found that participants did, indeed rank their alcohol consumption in terms of their own position in the rank of those around them, as opposed to using the average consumption as a benchmark. Indeed, the position in relation to alcohol consumption that participants believed they occupied in the ranking of other drinkers predicted the risk of alcohol related disease that they believed themselves to be at (Wood et al., *ibid*). This research showed how participants took into account all the comparison attribute values presented to them (in this case level of alcohol consumption) to give their own alcohol consumption a final ranking. This would suggest the potential influence of the quality of all fingerprints on an exhibit when a practitioner is making a comparative quality determination, and would suggest that the relative quality of each background mark in relation to the target mark would be a greater determination of target mark quality (the rank of the target mark) than the average quality of the background marks on the exhibit. This study has been used as supporting evidence to suggest that information should be presented differently in social norm interventions to work with the natural processing mechanism of the audience. It is important that the processing mechanisms of laboratory practitioners in relation to the effect of background marks quality are further empirically explored, so as to provide the correct information to influence mark submission policy and training, in order to improve reliability and efficiency of the process. Indeed, the likelihood of a specific target mark being submitted could be said to be predicted through the use of a decision by sampling model by ranking fingerprints on an exhibit against predetermined quality attributes and determining the positioning of the target mark along this scale. The threshold for submission would then determine whether or not the target mark would be submitted.

Table 6.1 - The application of decision by sampling theory to fingerprint submission (Adapted from Stewart et al. 2009)

Decision by sampling stage	Financial example (Stewart et al. 2009)	Potential application to fingerprint submission decision
STAGE 1: A target attribute is randomly selected	Either option a) or b) is selected, then a value or probability is selected, then one of the two values or probabilities is selected (e.g. a), value, £4,000)	Prospects in the submission choice would be A) choose to submit B) choose not to submit. Attributes of these prospects would be factors considered to determine quality within the mark e.g. minutiae count of 4 (clarity/ ridge flow examples of other possible quality attributes) is selected
STAGE 2: A comparison attribute is randomly selected from the decision sample (the decision sample comprises immediate decision context and long term memory – the use of either is equally likely)	A comparison attribute (a value of £2000) is selected from memory	A comparison attribute (e.g. a minutiae count of 8) is selected from the immediate context (neighbouring mark)
STAGE 3: The target and attribute comparison values are compared with a binary ordinal comparison	Target value of £4000 is compared to attribute comparison value of £2000. Target value is good in comparison.	Target value of 4 minutiae is compared to attribute comparison value of 8 minutiae. Target value is poor in comparison.
STAGE 4: If the comparison is favourable then the accumulator for the target prospect is increased by one count	One point is added to the accumulator for the target prospect	No point is added to the accumulator for the target prospect, a point is added to the accumulator for the alternative prospect
STAGE 5: If the difference between accumulator tallies for each prospect reaches a threshold, select the prospect with the highest accumulator. Otherwise, begin again at stage 1	If this point causes the difference between the accumulator tallies to cross a threshold then the target prospect (a) is selected.	If this point causes the difference between the accumulator tallies to cross a threshold then the alternative prospect (not submitting mark) is selected.

## 6.2. Summary of objectives

Research from within the domain of cognitive and social psychology has provided some insight into the potential influence that the background context can have on a number of decisions (including those of subjective quality). Consideration of both Decision by Sampling and the Inclusion and Exclusion Model provides two separate (but perhaps not mutually exclusive) examples of potential mechanisms for these effects. The ambiguous and subjective nature of the fingerprint submission decisions in relation to borderline fingerprints has been illustrated within Chapters 3-5 of this thesis, and, as such, it can be suggested that the background mark context of a target fingerprint may influence the quality assessment and submission of this target mark. The literature also suggests differences in the type of effects demonstrated according to sequential or simultaneous presentation of fingerprints. Given the common simultaneous presence of a background of additional fingerprints within fingerprint submission decisions this empirical study seeks to investigate the effect of the proportionate quality of background marks on the submission decisions made by laboratory practitioners in relation to borderline fingerprints when marks are presented simultaneously. Through doing so the study will meet the following objectives:

Objective 6.1. - An investigation of the main effects of background mark quality on target mark submission

- To empirically establish whether the proportionate quality of background (context) marks present on exhibits effects submission decisions made in relation to borderline (target) fingerprints presented within the same exhibit
- To determine whether assimilation or contrast effects occur within mark submission decisions made upon exhibits containing background marks, and whether the magnitude of the exemplar mark quality causes a change in any context effects identified
- To compare the effect of background mark quality on mark submission decision making within two UK laboratories

Objective 6.2. An assessment of individual differences in mark context effects

- To assess for individual differences in the effect of background mark quality context

Objective 6.3. Assessing for the influence of demographic factors and the presence of order effects

- To assess for any effects of order or positioning of exhibits and participant demographic factors on the effect of background mark quality context

## 6.3. Method

### 6.3.1. Summary of methodological process and requirements

In order to empirically assess the effect of the quality of background fingerprints present upon an exhibit on the submission decisions made in relation to ambiguous fingerprints present upon the same exhibit, it was essential to recreate the practitioner marking up task in a way that ensured a high level of ecological validity within a number of necessary experimental constraints. For example, it needed to be possible to complete the 'marking up' task remotely from the research team, and multiple copies of each exhibit needed to be created so that participants could be completing the task simultaneously, fitting participation around their casework demands. However, it also needed to be possible to follow standard marking up procedure and treat the experimental material in the same way as a physical exhibit. Furthermore, it was desirable that 'target' fingerprints (those for which submission decision outcomes were to be assessed) were of known usability according to the assessment of a fingerprint examiner in order to be able to assess submission in terms of 'ground truth'.

Experimental methodology involved the following five key stages:

- The selection of 'target' marks for use in the study
- The development and selection of 'context' fingerprints
- The creation of a series of mock exhibits with a range of proportions of good and poor quality context marks
- Laboratory practitioner participation in the study – 'marking up' exercise
- Data collection – determining the practitioner decisions made in relation to each target mark

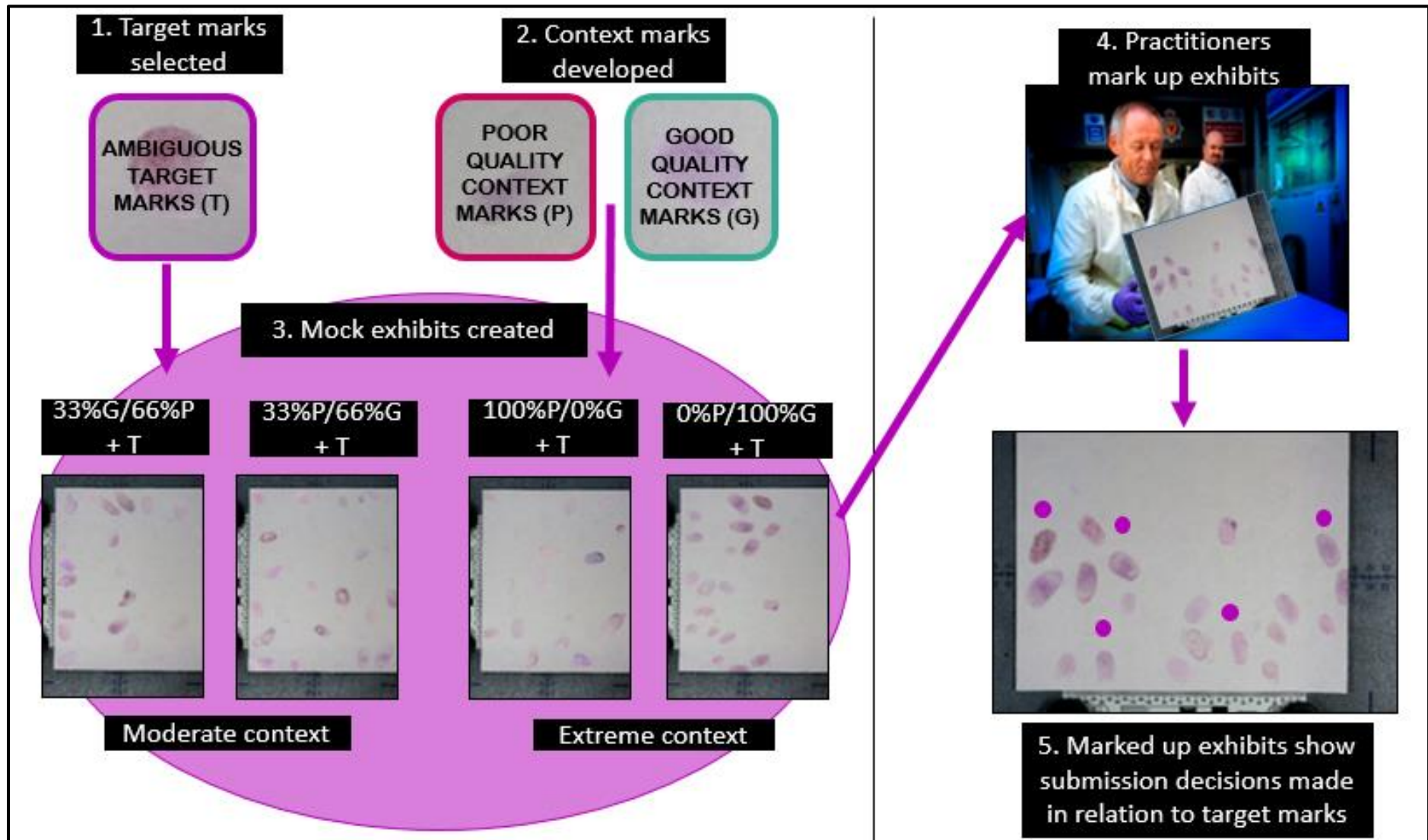
A high level summary of experimental methodology is provided in Figure 6.1.

### 6.3.2. Materials

#### 6.3.2.1. Selection of 'target' marks

In order to limit the requirements for operational fingerprint examiner participation in this study, target marks were taken from the series of examiner assessed fingerprints used within the experimental image set described within Chapter 3 and also utilised within Earwaker et al. (2015). All 20 of these 'borderline quality' fingerprints were selected to be used in the present study. Previous experimental work had demonstrated that these marks were ambiguous in relation to usability (from the view point of a laboratory practitioner), making them ideal for use as ambiguous target marks.

Figure 6.1- Summary of experimental methodology





### 6.3.2.2. The production of 'context' marks

#### Mark deposition and development

Three donors deposited a series of latent marks on clean sheets of white A4 Xerox Performer laser and inkjet printing paper (80g/m<sup>2</sup>). Each donor deposited marks across the surface of six sheets of paper. Varying deposition methodology was employed across the substrates in order to increase the likelihood of developing a combination of good quality and poor quality marks. Poor quality marks were deposited using light or swiping contacts with the substrate with the aim of achieving minimal deposition of ridge detail. Good quality mark deposition was achieved through the deposition of unloaded and natural secretion loaded depletion series using full fingerprint contact. Loaded deposition was made using a Latent Print Reference Pad: Amino Acid Based™ (Lightening Powder, part no. 1-2791) to load marks which were then deposited with moderate pressure in a number of depletion series across the substrate.

Approximately 24 hours after deposition the 18 sheets of substrate were treated with Ninhydrin, as per CAST recommended procedure for the visualisation of fingermarks on a porous paper substrate (Bowman, 1998). Visualisation took place at the CAST chemical visualisation laboratory at Sandridge, Hertfordshire. Ninhydrin working solution was made up as per CAST recommended formulation (Bowman, *ibid*) and applied to each sheet of paper in a fume cabinet. Treated papers were placed in a (Weiss Galenkamp) Ninhydrin Oven at 80°C and 75% relative humidity for 2 minutes.

Each sheet of treated paper was photographed in two sections using a Sony A77 camera on an Industria Fototechnica Firenze copystand with a 50 ml F2.8 macro lens. Papers were lit with Kaiser lights with daylight tubes. This ensured that all visualised marks were captured in a time efficient manner. The resultant 36 images were stored electronically for subsequent processing and printing.

#### Context mark selection

Images were viewed using Adobe Photoshop CS2 through a Toshiba Satellite laptop monitor. For each image (irrespective of intended deposition type) any developed ridge detail which appeared (to the author) to be of obviously very good quality was assigned a green coloured dot next to the mark on the Photoshop file, whilst any area of development which appeared to be of very poor quality (i.e. lacking detail or ridge development) was assigned a red colour dot. Consideration of developed ridge detail resulted in a total of 173 clearly poor quality areas of development and 72 clearly good quality marks.

Each photograph file was reopened and a new 'layer' was added to the image. Within this new layer any developed mark within the image were labelled with a code defining the donor or the mark,

whether it was to be included within the good quality or poor quality mark pool and a unique number (eg. AG1 (donor A, good quality, mark one)). Labelling these marks in this way meant that it would be possible to ensure the appropriate ratio of good and poor quality marks to each exhibit produced, and that a mixture of donors could be used across the exhibits. Marks were numbered in the order they were viewed in the photographs. An example of a labelled photograph file is provided in Figure 6.2.

Figure 6.2- Example of a labelled photograph file containing a mixture of good quality and poor quality marks



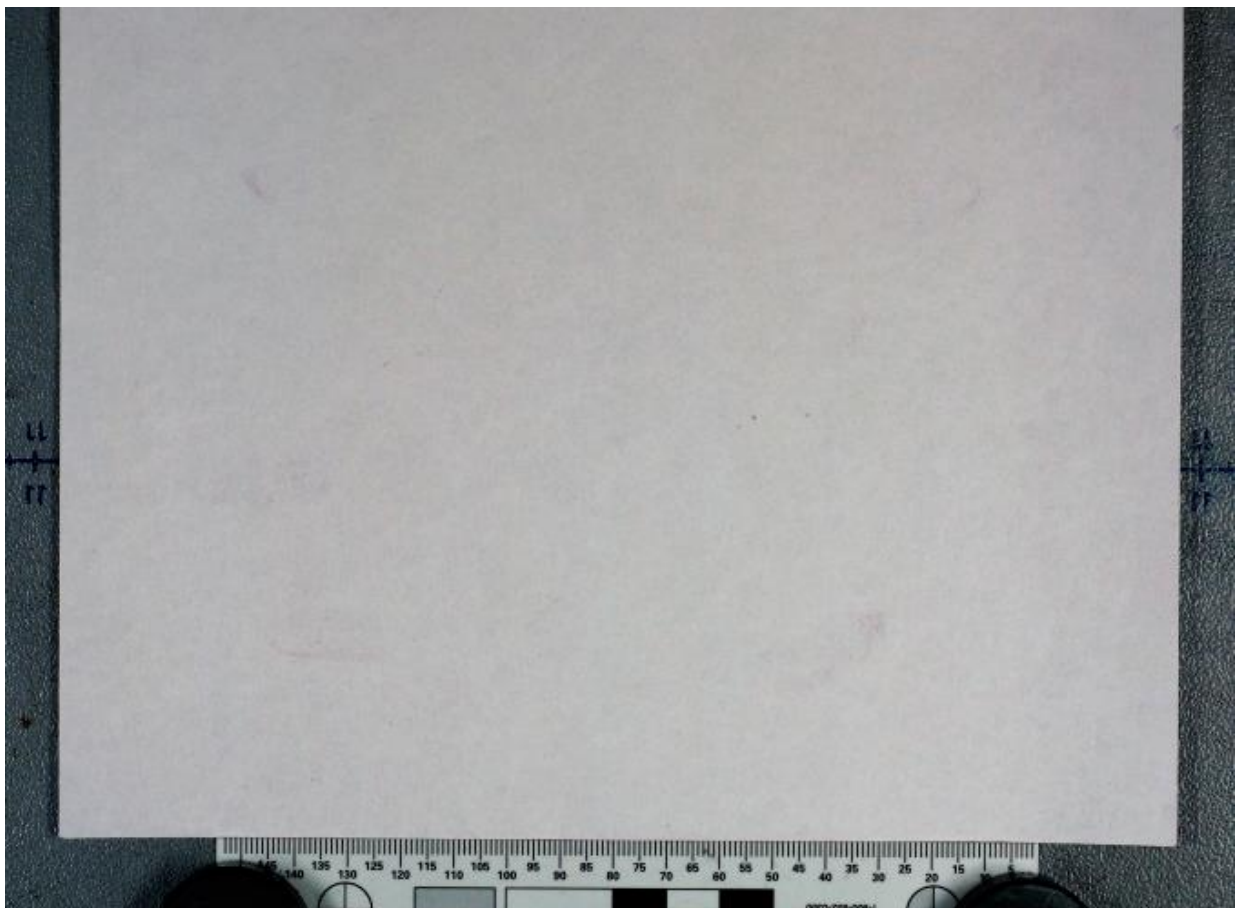
### 6.3.2.3. The production of mock exhibits using context marks

#### Selecting an exhibit background

In order to create a series of mock exhibits that would combine the ‘target’ marks which had been developed during previous experimental work (Earwaker et al., 2015), and the background marks developed for use in this study, consideration needed to be given to the most appropriate exhibit type and background upon which to present these marks. As all marks had been developed on paper using ninhydrin which often leaves a characteristic light purple or pink colour background stain, keeping the marks on such a background was considered to be a sensible approach in order to maintain continuity

across the exhibit. As such, one of the photographs taken was chosen to be used as the exhibit background as it did not contain any developed fingermarks upon the pink and purple coloured background. As the target images had been photographed with different imaging settings and a different colour profile it was necessary to darken the chosen image background so that the background colour would blend both with the selected 'target' marks and with the developed 'context' marks. The image was darkened by using the image adjustment settings in Photoshop CS2 to alter the colour balance. The background image is provided in Figure 6.3.

Figure 6.3- The background image used to create all mock exhibits



### Resizing 'context' images

In order to ensure a consistent appearance between target and context fingermarks it was important that images were presented at the same size ratio. Due to differences inherent in the two separate image capture systems pre-existing 'target' marks were smaller than the context images produced. In order to maintain the integrity of the target market images (upon which submission decision making was being judged), and due to reducing the size of an image being less damaging to its integrity than increasing its size (Marsh, 2015), images containing context fingermarks were scaled down to the same size as target marks.

## Exhibit creation

Figure 6.4 illustrates the composition of each of the 60 different mock exhibits created for use in the study.

Figure 6.4 - The content of mock exhibits

Experimental Session Details	Target Mark Position	Mock Exhibit Reference (Exhibits 1- 5)				
		Exhibit 1	Exhibit 2	Exhibit 3	Exhibit 4	Exhibit 5
<b>Quality Series 1</b> Good quality context marks: 0% (0) Bad quality context marks: 100% (16)	Position A	QS1A1	QS1A2	QS1A3	QS1A4	QS1A5
	Position B	QS1B1	QS1B2	QS1B3	QS1B4	QS1B5
	Position C	QS1C1	QS1C2	QS1C3	QS1C4	QS1C5
<b>Quality Series 2</b> Good quality context marks: 33% (5) Bad quality context marks: 66% (11)	Position A	QS2A1	QS2A2	QS2A3	QS2A4	QS2A5
	Position B	QS2B1	QS2B2	QS2B3	QS2B4	QS2B5
	Position C	QS2C1	QS2C2	QS2C3	QS2C4	QS2C5
<b>Quality Series 3</b> Good quality context marks: 66% (11) Bad quality context marks: 33% (5)	Position A	QS3A1	QS3A2	QS3A3	QS3A4	QS3A5
	Position B	QS3B1	QS3B2	QS3B3	QS3B4	QS3B5
	Position C	QS3C1	QS3C2	QS3C3	QS3C4	QS3C5
<b>Quality Series 4</b> Good quality context marks: 100% (16) Bad quality context marks: 0% (0)	Position A	QS4A1	QS4A2	QS4A3	QS4A4	QS4A5
	Position B	QS4B1	QS4B2	QS4B3	QS4B4	QS4B5
	Position C	QS4C1	QS4C2	QS4C3	QS4C4	QS4C5

The proportionate quality of context marks to be used in each exhibit set was referred to as a ‘Quality Series’ from 1 – 4 (labelled QS1-QS4). Within each quality series 5 mock exhibits were created each containing the same proportion of good and poor quality background marks (labelled A – E). Each of the exhibits, however, contained different (but equivalent quality level) marks. Repeating the same context or target marks within one ‘quality series’ was avoided as practitioners would be asked to examine all mock exhibits within the quality series in one instance, and it was intended that the exhibits should appear to be different to each other, as would be the case within normal casework. For each of the five mock exhibits created within each quality series, two additional, alternate versions were created (versions were labelled 1-3). Each version contained exactly the same context and target marks, but the positioning of the target marks was different in each. The position of the target marks was semi randomly chosen, giving three different semi-random position of target marks. This allowed counterbalancing to enable investigation of positioning effects within the study.

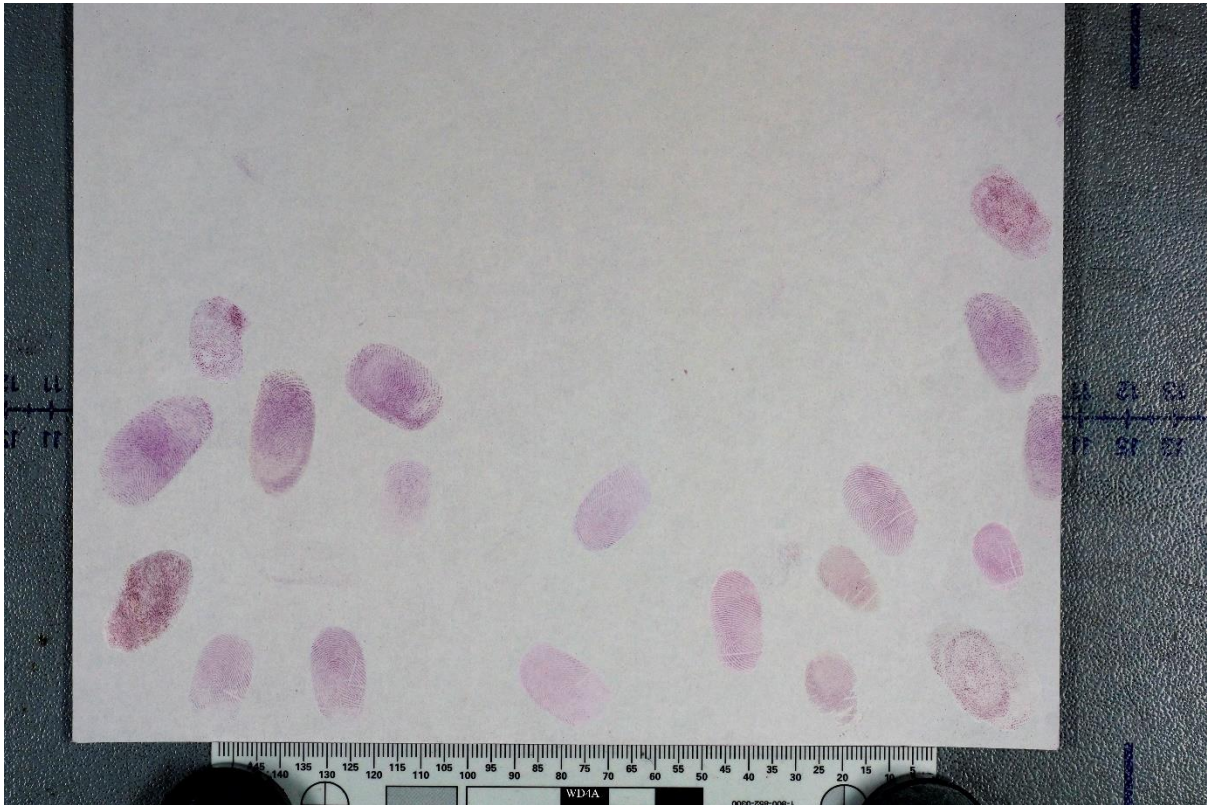
Context marks were selected for addition to mock exhibits so as to ensure that the correct proportion of good and poor quality marks were used in each case. Marks from different donors were used within each exhibit as it would be commonplace to encounter multiple donors within an exhibit from a crime scene. Context marks were added to exhibits in the order in which they were labelled within the mark and photographs taken until the determined number of marks of that quality rating and donor had been included. Marks were semi randomly distributed across the exhibit background, although on occasion similarity of background colour and tone dictated the most appropriate position for the mark in order to ensure a realistic exhibit was created.

Four 'borderline' target marks were included within each exhibit. Two of the target marks used in each case had been determined as 'usable' (suitable for identifying) and two had been determined as 'insufficient for comparison' by fingerprint examiners (see section 3.3.1.1 for further details). A different four target marks were used in each of the five exhibits included within each 'quality series'. Target marks were randomly assigned into pairs of the same usability and these pairs were randomly matched with a pair of the other usability determination giving five sets of four target marks. Each set was then randomly assigned to each of the 5 exhibits. The target marks were repeated in the same combination for each of the quality series. Inclusion of the same target marks in each quality series allowed a direct comparison of the effect of background mark quality on the submission decisions made in relation to the same borderline marks. However, this meant that it was important that the target marks should appear consistent with the exhibit they were on and did not stand out as being repeated or as being the focus of the study. To attempt to ensure this, different context marks were used for each exhibit and gaps of at least two weeks were left between the presentation of each quality series to practitioners.

Exhibits were created by copying the selected target marks and context marks using a hand drawn marquee tool from their original images and pasting them as image layers on top of the background image (provided in Figure 6.3). Using the hand drawn marquee tool enable images to be closely cropped from their background, lessening the appearance of contrast between the original background of the mark and the 'exhibit' background on to which it was pasted. Once pasted on to the background the colour levels of the mark were adjusted slightly in order to provide a good representation of homogeneity with the background.

An example of a mock exhibit is provided in Figure 6.5. Images of each mock exhibit detailing the origin and location of each mark are included in Appendix E.1. All final experimental exhibit images (as printed and used in the study) are provided in Appendix E.2.

Figure 6.5 - An example of a mock exhibit



#### Mock Exhibit Printing

Prior to printing, exhibit files were flattened to produce jpeg images (as opposed to multi-layered Photoshop image files). The exhibit images were also rescaled to ensure that the images could be printed 1:1 at A4 size. Final images were a resolution of 400 pixels per inch and 25.64cm by 17.09 cm in size.

The quality of the scaled images was appraised to ensure they were still of the appropriate quality. 10 colour copies of each of the 60 exhibit photographs were printed at a ratio of 1:1 within a UK police force photographic department on photographic paper, producing 600 A4 printed photographs. Printed images were utilised in this study in order to ensure consistency in the image resolution provided to participants. Had electronic images been provided to participants then these would have needed to have been viewed on monitors provided at the place of work of the participants. As police forces across the UK have differing facilities and equipment it is likely that participants would have viewed the images at on screens of differing resolutions, limiting the reproducibility of the task in each location.

#### 6.3.2.4. Exhibit pack compilation

Ordering and counterbalancing



A system of counterbalancing was employed within the experimental design in order to balance the data against, and to allow for the assessment of, the effect of:

- the position of target marks within an exhibit
- the order of presentation of the proportionate quality of context marks (quality series)

Counterbalancing was achieved by semi randomly allocating each participant to one of three groups in relation to the positioning of target marks within the exhibits received (A, B, or C), and to one of four groups (category 1, 2, 3, or 4) in relation to the order of presentation of different proportions of context mark quality (quality series). See Table 6.2, Table 6.3 and Table 6.4.

Table 6.2 - Key to counterbalancing groups for target mark positioning

Group	Meaning
<b>A</b>	Random distribution 'A' of target marks
<b>B</b>	Random distribution 'B' of target marks
<b>C</b>	Random distribution 'C' of target marks

Table 6.3 - Order of quality series for each category group

	Order of quality series (QS)			
	Quality Order 1	Quality Order 2	Quality Order 3	Quality Order 4
Exercise 1	QS1	QS4	QS3	QS2
Exercise 2	QS2	QS3	QS4	QS4
Exercise 3	QS3	QS2	QS1	QS1
Exercise 4	QS4	QS1	QS2	QS3

Table 6.4 - Key to make up of quality series

Key to quality series			
Quality Series	Percentage (and number) of:		Total number of marks (including 4 target marks)
	Good quality context marks	Poor quality context marks	
QS1	0% (0)	100% (16)	20
QS2	33% (5)	66% (11)	20
QS3	66% (11)	33% (5)	20
QS4	100% (16)	0% (0)	20

The allocation of both counterbalancing conditions to each participant is given in Table 6.5. Participants from both laboratories were allocated references beginning at participant reference A.

Table 6.5 - Counterbalancing conditions assigned to each participant

Participant Ref	Counterbalance reference	Order of quality series	Target mark location
<b>A</b>	Quality order 1, position A	1,2,3,4	A
<b>B</b>	Quality order 2, position A	4,3,2,1	A
<b>C</b>	Quality order 3, position A	3,1,4,2	A
<b>D</b>	Quality order 4, position A	2,4,1,3	A
<b>E</b>	Quality order 1, position B	1,2,3,4	B
<b>F</b>	Quality order 2, position B	4,3,2,1	B
<b>G</b>	Quality order 3, position B	3,1,4,2	B
<b>H</b>	Quality order 4, position B	2,4,1,3	B
<b>I</b>	Quality order 1, position C	1,2,3,4	C
<b>J</b>	Quality order 2, position C	4,3,2,1	C
<b>K</b>	Quality order 3, position C	3,1,4,2	C
<b>L</b>	Quality order 4, position C	2,4,1,3	C
<b>M</b>	Quality order 4, position B	2,4,1,3	B

The composition of experimental packs and distribution instructions

As previously stated, practical requirements and time restraints meant that the experimental exercise needed to be carried out remotely within the participant's places of work, without the direct input of



the researcher. In order to ensure consistency across participation methodology and conditions experimental paperwork was packaged and labelled in a way that would allow straightforward distribution of the materials by an appointed stakeholder within the participating laboratory.

Mock exhibits were collated according to the requirements of the experimental design and counterbalancing, as set out in Tables 6.2 – 6.5, and were placed in four A4 envelopes to be provided to participants.

Experimental materials provided to participating laboratories

The appointed research coordinator (a senior stakeholder within each of the two participating laboratories) was sent a pack of experimental materials and full instructions for the distribution of these materials. The pack contained:

- Written instructions for the distribution of experimental packs (see Appendix E.3)
- Record of pack distribution indicating which pack to be given to each participant at each time (see Appendix E.4)
- 4 envelopes (Quality series 1, 2, 3, and 4) containing experimental materials for each of 12 practitioners
- Envelope contents:
  - First time participation envelopes (white A4) labelled e.g. C (context),A (participant reference) 1 (time 1) containing:
    - Instruction sheet (Appendix E.5)
    - Participant information sheet (Appendix E.6)
    - Informed consent form (Appendix E.7)
    - Demographic information form (Appendix E.8)
    - Participant instructions (Appendix E.9) attached to a set of five mock exhibits (hole punched at top edge and attached with treasury tags)
    - Set of purple dot stickers (equivalent to highest possible number of marks)
  - Time 2, 3, and 4 envelopes (white A4) labelled using same method containing:
    - Participant instruction sheet
    - Instructions attached to set of five mock exhibits (hole punched at top edge and attached with treasury tags)
    - Set of purple dot stickers (equivalent to highest possible number of marks)
- Verbal instructions to be provided to participants (Appendix E.10)

### 6.3.3. Participants

Participants were, initially (25) fingerprint enhancement laboratory practitioners. 13 practitioners were employed within Fingermark Enhancement Laboratory A and 12 practitioners were employed by Fingerprint Enhancement Laboratory B. Practitioners had an average of 9.5 years' experience working within a laboratory (with a minimum and maximum of 1 year and 20 years, respectively). All practitioners had received initial training from the College of Policing and in house continued training. The participants recruited were all the available laboratory practitioners from Laboratories A and B (rather than being a selected sample from each laboratory). These were the same practitioners who also participated in the study outlined in Chapter 5. Participation in the present study was carried out prior to participation in the minutiae count study. Participation was ordered in this way to prevent participants from studying the target marks (also used as part of the mark image set in Chapter 5) too closely prior to carrying out the marking up task detailed within this chapter. Participation in the two studies was split to ensure that there was at least a two week gap between participants completing the final exercise of the present study and the first activity of the minutiae count study described in Chapter 5. This was deemed to be of sufficient time to allow participants to have viewed many casework marks in between the two activities and thus to remove any order effect relating to having previously participated in another study.

#### 6.3.4. Procedure

Research coordinators distributed experimental packs amongst laboratory practitioner participants as per the distribution instructions provided (see Appendix E.3).

Within the packs participants were provided with written instructions detailing how to complete the experimental task. Participants were instructed to:

“Look at the first of the exhibit images provided. Mark up this exhibit, labelling all marks which are sufficient for submission to the bureau by placing a purple sticker next to each of these marks. Please draw an arrow on the sticker to clearly indicate which mark the sticker relates to. Carry out this task at your normal workplace, using your usual marking up procedures and equipment. Spend the same amount of time considering the exhibit that you would typically in normal casework. Repeat this process for all 5 exhibit photographs, ensuring that you work through them in the order given.”

Full instructions provided can be seen in Appendix E.5. In addition to the written instructions, brief verbal instructions were provided to each participant by the research coordinator. The verbal instructions were read out to each participant on each participation occasion and stated that participants should carry out the marking up task as per the written instructions provided, assuming that the exhibits had been submitted as part of a volume crime case (see Appendix E.10).

Coordinators were instructed to administer each of the sets of exhibits to practitioners with at least a two week gap between each experimental session. This element of the experimental design was intended to ensure that the repeated use of the same target marks across each experimental repeat was not detected by practitioners and to lower the chance of any familiarity or practice effects associated with making more than one submission decision in relation to these marks.

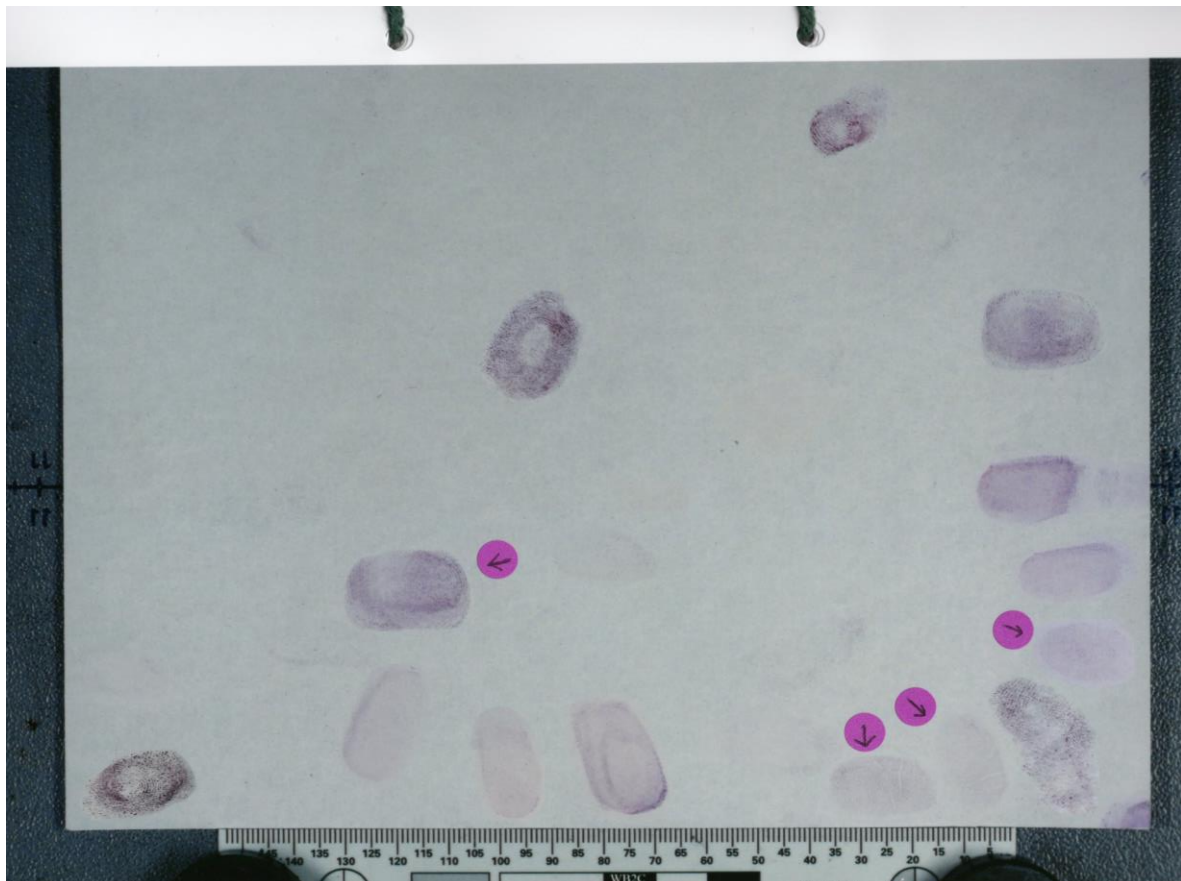
Upon completion of each experimental task, practitioners were instructed to return their completed and sealed envelope containing their marked up mock exhibits. All responses were stored by the coordinator until the experiment had been completed at which point the packs were collected by the researcher for analysis.

## 6.4. Results

### 6.4.1. Preparation and coding of data

Experimental data in the form of 'marked up' exhibits was collected from research coordinators of both laboratories. An example of a 'marked up' exhibit Figure 6.6.

Figure 6.6 - Example of a marked up exhibit



In order to analyse this data the marked up exhibits were each coded in order to quantify the responses given. During this process the following information was gleaned and recorded in relation to each participant:

- The laboratory within which the practitioner was employed
- The participant reference assigned to the practitioner carrying out the exercise
- The gender of the participant
- The age of the participant
- The job title of the participant
- The years of experience of the participant
- The order of context quality conditions assigned to the participant
- The position of target marks assigned to the participant

The following information was gleaned from each mock exhibit exercise completed by each participant:

- The number of target marks 'marked up' (selected for submission)
- The number of context marks 'marked up' (selected for submission)

Tabulated raw data for each participant is included in Appendix E.11.

#### 6.4.2. Summary of demographic information

Of 25 initial participants, a total of 24 laboratory practitioners completed the full study. One practitioner began the study but failed to complete all four experimental exercises. Because of the incomplete data for this participant and the potential for this to skew the data set given the relatively small total sample size it was decided that the responses provided by this participant should be removed from the data set prior to analysis.

Of the 24 practitioners who fully completed the study, 12 were employed by Laboratory A and 12 were employed by Laboratory B. 41.67% (10) of participants were male and 58.33% (14) were female. All participants were employed as Fingermark Enhancement Officers (either termed FLO or FASLO, dependent upon the accepted terminology of the organisation). Practitioners had a mean experience of 9.5 years, and a range of experience from 1.6 to 20 years.

#### 6.4.3. Objective 6.1. The main effect of context (background mark quality)

Assessment of interaction effects (effect of laboratory on background mark effect)

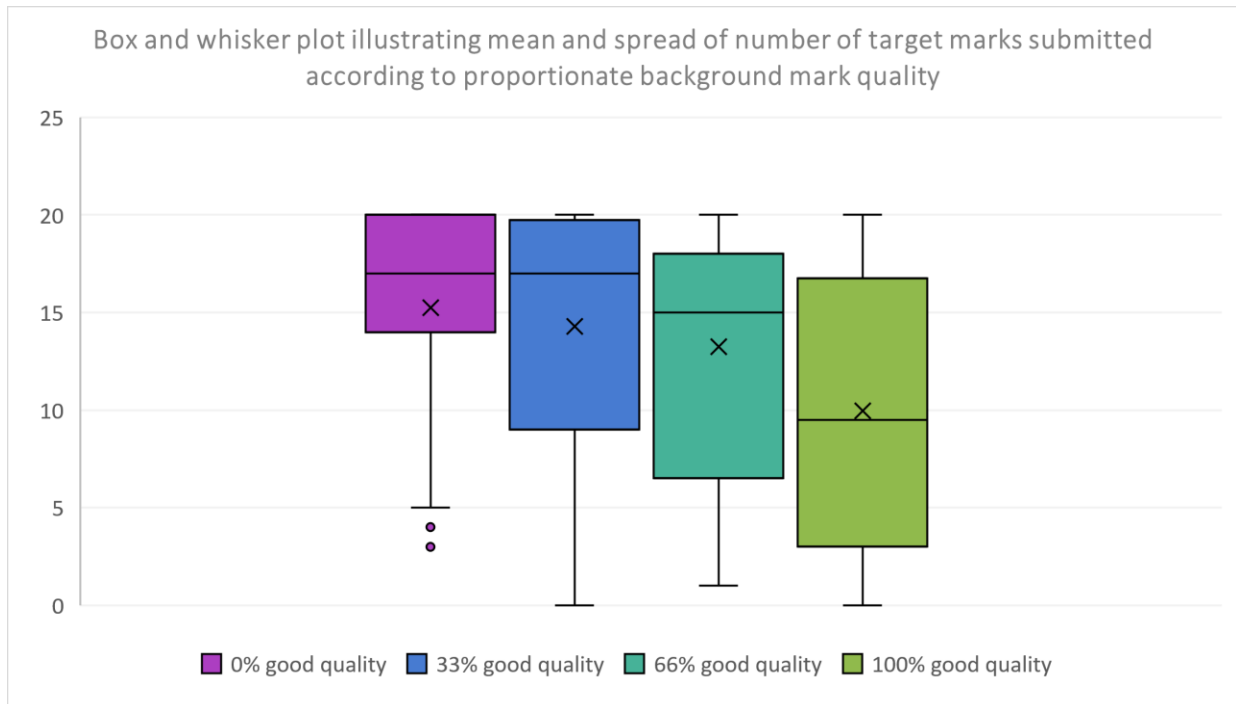
Descriptive statistics (provided in Table 6.6) show some slight differences in target mark submission between the two laboratories. Mean target mark submission is slightly higher within the Laboratory B from 'Quality Series 2' onwards. The spread of the mean number of target marks submitted is consistently greater within the Laboratory A.

Table 6.6 - Descriptive statistics for numbers of target marks submitted for each laboratory

Proportion of background mark quality	Laboratory A			Laboratory B			All Practitioners
	n	M	SD	n	M	SD	Total number of target marks submitted
0% good quality ('Quality series 1')	12	15.50	6.08	12	15.00	4.60	366
33% good & 66% poor quality ('Quality series 2')	12	13.67	6.96	12	14.92	5.79	343
66% good & 66% poor quality ('Quality series 3')	12	13.00	6.13	12	13.50	6.01	318
100% good quality ('Quality series 4')	12	9.75	8.16	12	10.17	6.38	239

Differences in the mean and spread of number of target marks submitted according to proportionate background mark quality are presented in Figure 6.7. This box plot illustrates the spread of data in relation to target mark submission according to the proportion of background mark quality. This demonstrates the decrease in the mean numbers of target marks submitted as the percentage of good quality background marks increased.

Figure 6.7 - Box and whisker plots illustrating spread of number of target marks submitted according to proportionate mark quality



Statistical analysis was initially carried out to determine whether there were statistical differences or interactions between the laboratory in which practitioners worked and the effect of background mark context on target mark submission. The data was found to meet the necessary assumptions of normality, homogeneity of variance, and sphericity, thus a mixed between-within-subjects analysis of variance (4 x 2 ANOVA) was carried out to assess the impact of laboratory (Lab A, Lab B) on the number of target marks submitted by practitioners, across the four levels of background mark quality (0% good quality, 33% good quality & 66% poor quality, 66% good quality % 33% poor quality, 100% good quality). There was no significant interaction between laboratory and level of background mark quality, Wilks Lambda = 0.963,  $F(3, 20) = 0.255$ ,  $p=0.857$ , partial eta squared = 0.037, suggesting no difference in the effect of background mark quality on target mark submission between the two laboratories. There was a substantial main effect for the proportion of background mark quality, Wilks Lambda = 0.503,  $F(3, 20) = 6.586$ ,  $p=0.003$ , partial eta squared = 0.497, with both groups showing a reduction in the number of target marks submitted as the proportionate quality of the background marks increased (see Table 6.6). The main effect comparing the two laboratories was not significant,  $F(1, 22) = 0.08$ ,  $p=0.779$ , partial eta squared = 0.004, suggesting no overall difference in mark submission between the two laboratories.

Pairwise comparisons

Pairwise comparisons were made within the one-way repeated measures ANOVA to determine where there were statistically significant differences between the numbers of target marks submitted according to the proportionate quality of context marks. A summary of the output of these comparisons is provided in Table 6.7.

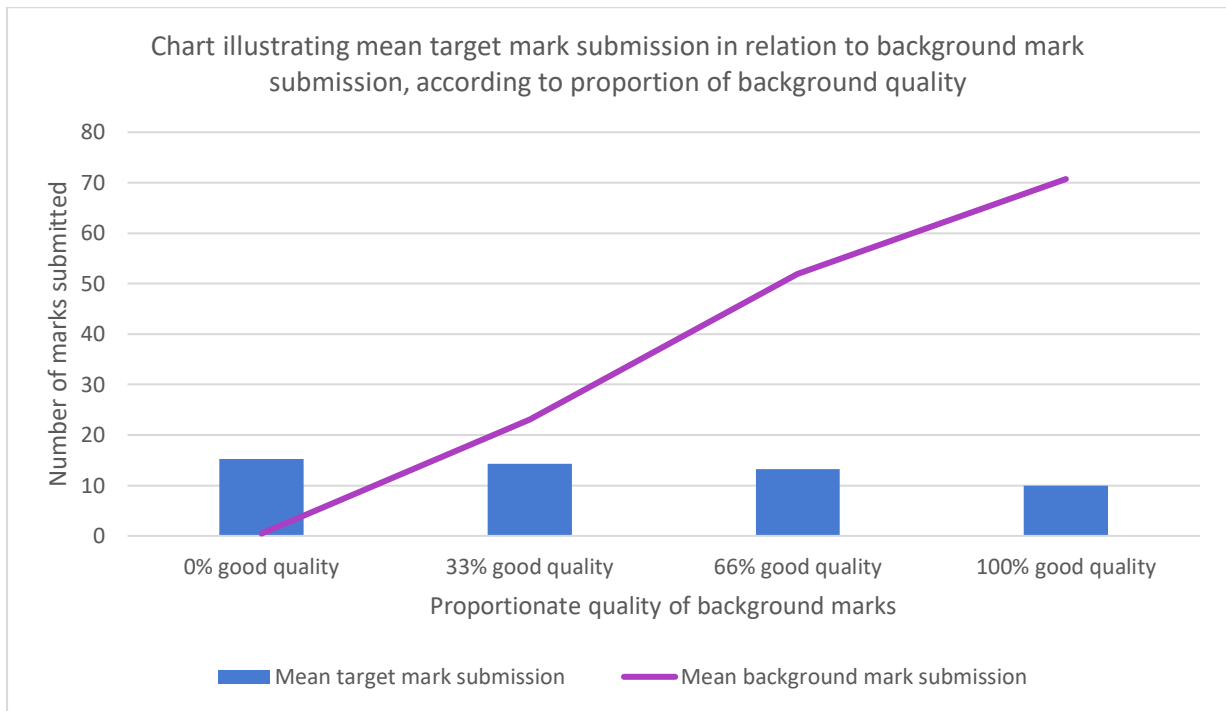
Table 6.7 - Pairwise comparisons testing for significant differences

Comparison between:	P value	Statistically significant difference?
0% good quality and 33% good quality/66% poor quality	0.766	No
0% good quality and 66% good quality/33% poor quality	0.074	No
0% good quality and 100% good quality	0.001	Yes
33% good quality/66% poor quality and 66% good quality/33% poor quality	0.360	No
33% good quality/66% poor quality and 100% good quality	0.001	Yes
66% good quality/33% poor quality and 100% good quality	0.015	Yes

Pairwise comparisons show statistically significant differences between the number of borderline target marks submitted by practitioners when the background context of the mock exhibit was 100% good quality (0% poor quality) marks and each of the other three proportions of background quality.

Figure 6.8 illustrates a comparison of target mark and context (background) mark submission according to the percentage quality make up of background mark quality. As the proportionate quality of the background marks increases the mean number of background marks submitted increases and the mean number of target marks submitted decreases. This provides a manipulation check in relation to the selection of fingermarks for used as background marks in the study. The number of background marks submitted has increased as the proportionate intended quality of those fingermarks increased, thus validating the procedure used.

Figure 6.8 - Relationship between target and background mark submission



#### 6.4.4. Objective 6.2: Individual differences in mark submission

In order to further explore the relationship between background mark quality and borderline mark submission, target mark submission was divided according to practitioner. The division of the data in this way allows consideration of individual differences and variation in mark submission levels as well as differences in the effect of background mark quality on submission.

Figure 6.9 illustrates the variation in target mark submission according to the proportionate quality of the background marks present for each of the 24 practitioners. Practitioners from Laboratory A are labelled 1A – 1L and practitioners from Laboratory B are labelled 2A- 2L. In the schematic, participants are divided according to whether they demonstrated an overall assimilation effect, no overall change in mark submission according to background mark context, or a contrast effect. Indeed, this schematic demonstrates there to be differences in the magnitude of context effect shown between practitioners. Some practitioners show a considerable contrast effect in their target mark submission (for example 1B and 1I), whereas some practitioners exhibit a slight overall assimilation effect (for example 1F). The schematic also illustrates substantial individual differences in the general quantities of target marks submitted. For example, practitioner 1C consistently submits very low numbers of the target marks,



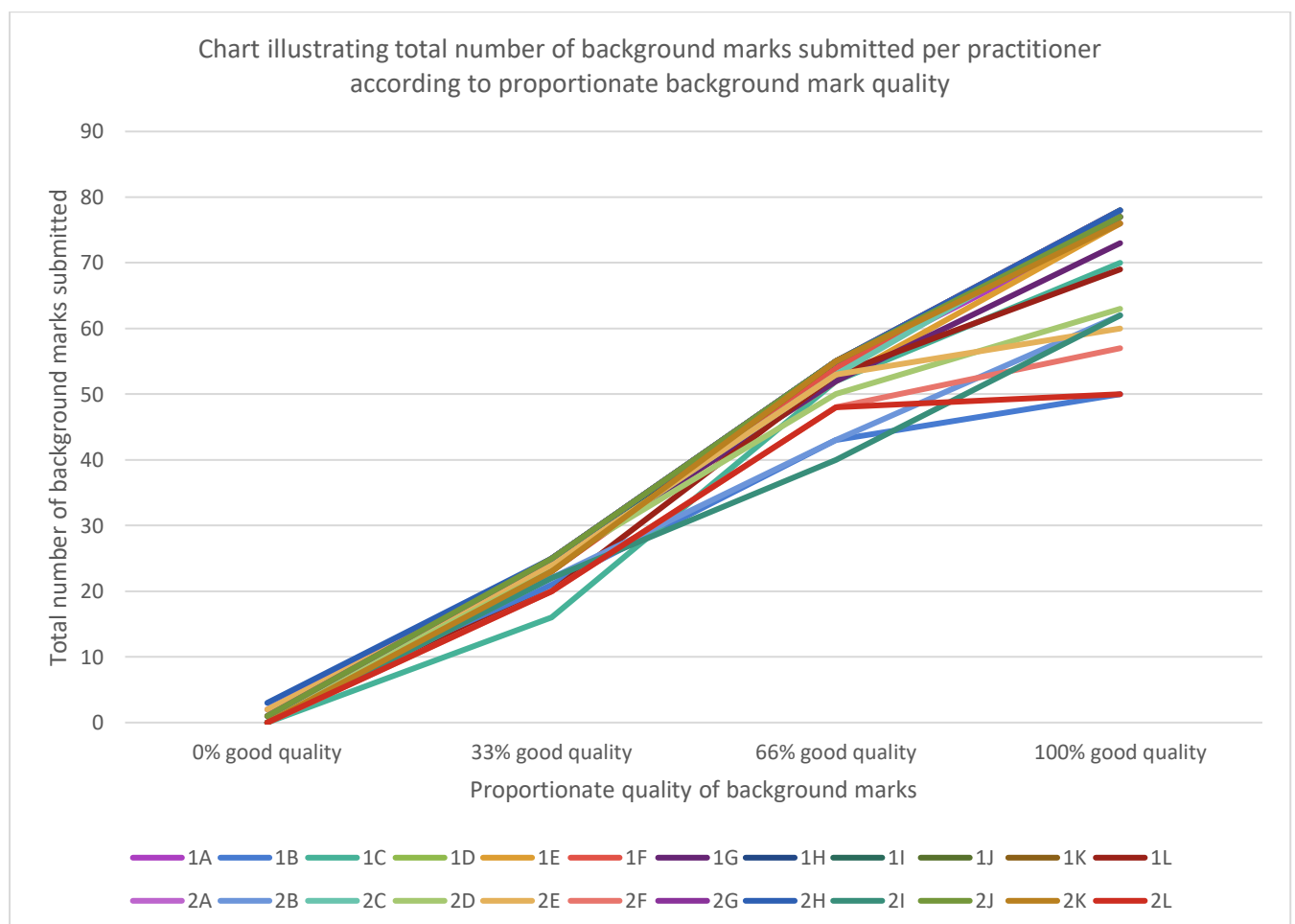
Figure 6.9 - Total number of marks submitted by each practitioner according to context mark quality



whereas practitioner 2H submits all target marks in all cases, illustrating a ceiling effect. This suggests a range of thresholds for submission in practitioners in relation to these borderline marks.

In order to further explore this individual difference in relation to the background marks the total number of background marks was plotted for each participant (see Figure 6.10). Figure 6.10 illustrates less individual difference in mark submission in relation to the non-ambiguous background marks than was illustrated in Figure 6.9 in relation to the ambiguous target marks with all practitioners submitting more fingermarks as the proportionate quality of the background fingermarks increased. This relationship again provides validation for the procedure employed. There can be seen to be greater individual differences in the number of marks submitted as the proportion of good quality marks increases. This would suggest that there may have been slightly more variation in the quality of the 'good quality' mark set than in the 'poor quality' mark set. This is an unsurprising finding as poor quality marks tended to lack any developed ridge detail whereas there were varying degrees of detail present within the good quality marks.

Figure 6.10 - Total number of background marks submitted per participant



#### 6.4.5. Objective 6.3: Assessing for the influence of demographic factors and the presence of order effects

##### 6.4.5.1. An investigation of order and position effects

The design of this experiment included counterbalancing to limit, and enable an assessment of, any effects of the order in which participants completed marking up tasks in relation to exhibits with different proportionate context mark quality, and any effects of the positioning of target marks within the exhibit.

##### Analysis of order effects

A between-within-subjects ANOVA was carried out to determine whether the order of presentation of the proportionate background mark quality of the mock exhibits had an influence upon the number target of marks submitted within each type of background quality. Analysis found that there was the same change in submission of target marks according to background proportionality for all four orders of background mark quality presentation (no interaction effect was found between the two variables (Wilks Lambda = 0.438,  $F(9, 43.96) = 1.975$   $p=0.066$ )).

Testing for between-subjects effect found no significant difference in target mark submission between the four different orders of background quality presentation ( $p=0.632$ , partial eta squared = 0.081).

##### Analysis of position effects

A between-within-subjects ANOVA was carried out to determine whether the position of the target marks on the mock exhibits had an influence upon the number of target marks submitted within each target mark position. Analysis found that there was the same change in submission of target marks according to the position for all three positions of target marks (A, B, and C) (no interaction effect was found between the two variables (Wilks Lambda = 0.887,  $F(6, 38) = 0.392$   $p=0.88$ )).

Testing for between-subjects effect found no significant difference in target mark submission according to the position of target marks ( $p=0.230$ , partial eta squared = 0.131).

##### 6.4.5.2. Influences of demographic factors on effect of context mark quality

Analyses were carried out to further explore the effect of context mark quality in relation to a number of participant demographic factors. Demographic information collected through the 'demographic information sheet' completed by each participant was used to assess for any interaction and between-subjects effects on the effect of proportionate background mark quality on target mark submission according to gender and years of experience as a fingerprint enhancement officer.

No significant interaction or between-subjects effect was found according to these demographic factors. A summary of statistical outputs is provided in Table 6.8.

Table 6.8 - Tabulated results of mixed between-within ANOVA analyses testing for interaction and between subject effects according to key demographic factors

Demographic factor	Data grouping	Test statistics for interaction effect		Test statistics for between subjects effect	
		Wilk's lambda	p	p	Partial eta squared
Gender	1= male 2= female	0.853 (3, 20)	0.352	0.347	0.040
Experience	1 = < 10 years 2 = ≥ 10 years	0.902 (3, 20)	0.548	0.852	0.002

### 6.5. Summary of key findings

The key findings of empirical research within this chapter are summarised as follows:

Objective 6.1. - An investigation of the main effects of background mark quality on target mark submission

- There was a decrease in the mean number of target marks submitted as the proportionate quality of the background marks present increased
- There was a significant difference between the number of target marks submitted when exhibits which contained an extreme good quality background context (100% good quality marks) and each of the other proportions of background mark quality
- There was no difference between the effect of the proportionate background mark quality on target mark submission between the two laboratories

Objective 6.2. An assessment of individual differences in mark context effects

- There are considerable individual differences in the number of target marks submitted across the experimental conditions
- There is some individual difference in the overall directionality of effect of context mark quality on borderline mark submission (individual differences in assimilation and contrast effects)

Objective 6.3. Assessing for the influence of demographic factors and the presence of order effects

- No exhibit order or target mark positioning effects were found
- There was no significant difference in context effect according to gender or years of experience within the laboratory

## 6.6. Discussion

### 6.6.1. Objective 6.1: An investigation of the main effects of background mark quality on target mark submission

There was an overall decrease in the submission of borderline (target) marks as the proportionate quality of background (context) marks increased. This finding is supported by the theory of decision by sampling (Stewart et al., 2006) and is also an example of a contrast (as opposed to an assimilation) effect based upon the inclusion/exclusion model (Bless & Schwarz, 2010).

Decision by sampling suggests that the rank of the quality of a target mark within the decision context will determine the perceived quality of the mark (Stewart, 2009). In the present study, the decision context of each exhibit was engineered to provide two distinct types of fingerprint: those of very good quality (clear marks with good definition between ridges and furrows, containing a large number of visible minutiae) and those of very poor quality (marks with poor definition and contrast with no, or very few visible minutiae). This should mean that a series of comparative quality determinations would, in each case, result in the borderline target marks ranked above each of the poor quality marks and below each of the good quality marks.

This theoretical ranking matches the trend in the data for higher quality determination of target marks (leading to a higher number of target marks submitted), the higher the number of poor quality marks within the background context. However, there was only a statistically significant change in target mark submission when extreme good background context was presented and target mark submission was compared with each other group.

The presence of such an overall relationship between background mark quality and target mark submission suggests that the practitioners may, indeed, have been utilising the decision background to make comparative quality judgements, as opposed to making these comparisons from information within their working memory. Decision by sampling requires the application of a threshold (cut off point) upon which it is considered that either the exemplar or target has been determined to 'win' after a series of pairwise comparisons. It should also be born in mind that each exhibit contained four different borderline target marks between which the outcomes of a comparison would be much more challenging and uncertain. The perceived relative quality of other target marks present would have had the potential to affect the ranked position of the target mark in question slightly.

A further consideration in relation to the application of Decision by Sampling Theory is the selection of a target attribute by practitioners during their series of comparative judgements between the target mark and the context marks. Previous research within Chapter 4 of this thesis has suggested that

quantity of ridge detail and clarity of mark are important factors in mark quality assessment and fingerprint submission decision making. So it may well be the case that these quality indicators are commonly employed as an attribute for comparison. However, it may instead be the case that a more holistic view of quality is considered. The theory of decision by sampling (and, indeed, wider aspects of judgment and decision making) may present opportunities to be used as a tool for further investigating the mechanism of the fingerprint quality determination process through isolating the quality attribute that has the greatest effect on mark submission. Further empirical study manipulating the quality features present in target and background marks could potentially be used to achieve this, through assessing the strength of the relationship between marks with certain quality indicators in comparison to target marks. For example, using marks which all have the same number of minutiae as both target and context marks, but have varying degree of clarity in one task, and using marks of similar clarity but with different numbers of minutiae in another case, and then comparing the effect of the proportionate quality of these background marks on target mark submission between the two cases, would be a potential methodology for achieving this aim.

The main results of this study can also be considered in terms of the inclusion/exclusion model. The terminology of this model would describe the overall effect of submitting more borderline target marks as the quality of the background marks decreased, as a contrast effect. According to the inclusion/exclusion model this would mean that the practitioner decision makers have utilised the context of the target mark (the quality of the surrounding marks on the exhibit) in order to inform their representation of a comparison standard, resulting in the contrast effect. For example, when the context of the target mark is good quality background marks this background contextual information is used as the comparison standard resulting in the target mark appearing to be of relatively poorer quality. Evidence for such an approach can, indeed, be found in the data, with the increases of ambiguous target mark submission as the average quality of the background marks decreases (as shown in Table 6.6). This may suggest that the practitioners view comparative quality of other marks as relevant in the decision process (Bless & Schwarz, 2010), arguably due to the highly subjective nature of the decision which may cause a reliance on an external point of comparison, instead of consideration of the target mark in isolation (as would have been suggested had an assimilation effect been shown to occur).

This empirical study was not designed or intended to identify or discriminate between decision mechanisms, instead it was intended to identify the effect of context mark quality of ambiguous mark submission. As such, it is not possible to use the present data to provide support for either the use of decision by sampling or the inclusion/exclusion model. Further research could use the present study as a model for manipulating the values of mean and pairwise comparisons between background mark

and target mark quality in order to provide support for a particular decision mechanism. However, such a study would require either an accurate knowledge of practitioner quality determination criteria, or would need to take a less naturalistic approach and be prescriptive about the quality criteria used in the task, in order to ensure that any inference could be made from the outcome which may reduce the ecological validity of the findings.

#### Operational significance of context effects

The presence of a contrast effect as demonstrated in this study, in an operational context may lead to the loss of identifiable marks in cases in which there are good quality background marks on the exhibit, and may lead to wasted resources in terms of poor quality marks submitted when dealing with exhibits made up of poor quality background marks.

Operationally, intentionally prioritising the examination of the best quality marks and having a relative threshold for mark submission (i.e. a 'take the best' policy) may well be a cost effective approach as many scientific support departments struggle to process a large volume of evidence due to budget cuts and decreased staffing levels, in some cases as a result of mergers between fingerprint services of different police forces. It would seem sensible to prioritise marks that have a higher evidential value in this way. However, this needs to be a transparent and conscious process based upon a written and validated policy in order to ensure the credibility of the mark recovery process. If it is the case that practitioners are attempting to base their submission decisions on each mark in isolation without consideration of the overall quality of marks within the same exhibit, then the effect demonstrated within this study may be of concern as it would appear that practitioners are, in practice, taking the quality of background marks present in to consideration during the decision making process. Acknowledging such a subconscious effect could assist in quality assurance procedures, as well as encouraging practitioners to be aware of the surrounding qualities of a mark and how these may influence their submission decision.

An important possible extension of this work would be to determine the conscious or subconscious nature of this context effect by asking (at the time of carrying out the activity, as per best practice research into metacognition (Newell & Shanks, 2014)) participants whether they are making a conscious relative quality judgement (for example, a possible response for making a conscious judgement may be "because there were so many good marks I decided against sending these borderline marks even though I think they may be of good enough quality to use"). It would seem unlikely that this is a conscious approach in this case, especially for practitioners from within the Laboratory A who have a policy of sending only the 20 best marks on an exhibit (see 3.4.2.). Whilst this sounds as if it may lead to an increase in the likelihood of a comparative approach to quality

determination (which, indeed, it may) the results of this particular experiment do not show this, as each exhibit only contained 20 marks, suggesting that the practitioners from this lab would have sent all marks deemed of high enough quality, only resorting to a conscious comparative approach once they had realised that there were more than 20 marks present upon the exhibit. This may suggest that the context effect occurring here is subconscious. Equally there was no difference in the effect found between the two different laboratories suggesting that this difference in policy has not have an effect on the submission decisions made in this case.

The wording of the question in this experiment (the task instructions) may have had the potential to impact upon the level of submission of target marks in this study. Participants were asked to mark up each exhibit as they would in standard casework. As such they were not directly told to mark up all exhibits that were of sufficient quality for comparison. Phrasing the task in this way allows for the naturalistic inclusion of policy and practice adopted in casework in the experimental findings. Whilst this could have the potential to confound the finding of the study to some extent (for example, if a policy was to only submit ten marks from each exhibit then a ceiling effect would be reached as the exhibits used in the study contained twenty marks each), taking this approach meant that any effects of the proportionate quality of background marks are representative of those that would occur within real life casework which would follow any such policies. Indeed, discussion with the managers of both participating laboratories ensured that no such ceiling effect would be reached with a total of twenty fingermarks present on the mock exhibits.

The initial intention during experimental design was to avoid any mention of crime type (either serious or volume) in the task instructions of the study as it was thought that this information may bias the decision making outcome of the participants, as contextual information has been shown to influence subjective decision making within fingerprinting and other forensic domains (Earwaker et al., 2015, Dror et al., 2006, Nakhaeizadeh et al., 2014). By omitting such information, it was intended that participants would rely solely on the usability of the ridge detail itself when considering whether or not to send it to an examiner. Upon initial receipt of the experimental packs participants at one police laboratory (which later withdrew from the study due to a high casework load) asked, via the experiment coordinator, what type of crime the exhibits related to. The difference in submission policy and procedure between the two crime categories was discussed further with the laboratory manager in order to assess the potential impact of selecting each category of crime upon the experimental results. When asked, the laboratory manager stated that there was no difference between the policy and procedure for mark submission according to crime category, but that the laboratory practitioners felt that they could not make their decision without this information. This suggests that part of the practitioner submission decision relies upon crime type, even though according to the policy of the



laboratory the decision outcome should not differ for identical marks presented in each context. It was decided, in order to prevent confusion amongst practitioners and to keep the experimental conditions consistent across participating laboratories, that additional verbal instructions should be read to each practitioner providing them with information that they should treat the exercise as if marking up exhibits from a case of volume crime. This was the only reference that was made to crime type within the experimental materials in order to prevent it from becoming a point of focus within the experimental practitioner decision making task.

The practitioner requirement for crime type information without any procedural need for such information is an interesting incidental finding of this research. From the point of view of the present study the volume crime context may have had some influence upon mark submission levels throughout the experimental tasks. Previous research described within Chapter 3 of this thesis (Earwaker et al., 2015) found that type of crime context information provided in a marking up task effected the quality threshold for the submission of borderline marks. The threshold for mark submission within this laboratory increased when submitting experimental marks within the context of a volume crime resulting in a lower rate of submission of borderline quality marks. This may suggest that there was overall a higher threshold for mark submission in the present study than there would have been had the participants been asked to treat the exhibits as originating from a serious crime. However, Chapter 3 of this thesis demonstrated that a change in submission threshold according to crime context did not occur within a laboratory that routinely dealt with both types of crime, suggesting that in the present two laboratories (who both dealt with both types of crime) the crime context provided may not have had an effect. It would be interesting to compare the effect of the proportionate quality of background fingermarks on the submission of target marks presented in a serious and volume crime context through further research. Consideration was given to including these variables within the initial task, but it was decided that the focus of the task should remain on the quality assessment of the marks themselves, and that it was not practically viable to add the necessary counterbalancing conditions to the study to allow inclusion of the additional variable, given the relatively small number of participants and operational limits on the amount of time each participant could spend completing experimental tasks. As, such this remains a possible extension of the study for future work.

#### Comparison of contrast effects between laboratories

No significant differences in the context effects shown were found between the two police force laboratories (A and B) that participated in this study. This is not a surprising finding as the laboratories were both of a similar size with a similar range of routine casework, and similar make up of practitioners in terms of background and experience. In addition, both laboratories use the College of

Policing for practitioner initial training. The study was designed to limit the impact of a 'take the 20 best marks policy' held within Laboratory A by including only 20 marks in each mock exhibit. The similarities between the two forces would suggest that this policy has not affected the mark quality assessment and submission decisions made by practitioners when dealing with smaller numbers of marks.

#### 6.6.2. Objective 6.2: Individual differences in mark submission

Decision by sampling theory suggests that a series of pair wise comparisons are made between target and context stimulus attributes (Stewart, 2009). Empirical studies included in Chapter 4 of this thesis demonstrated that one of the most commonly used attributes to determine quality in a mark is minutiae count. Chapter 5 of this thesis provides empirical evidence to suggest that there is a considerable inter-practitioner variation in minutiae count in relation to the same fingerprint. This may explain why there are individual differences in numbers of marks submitted, if practitioners are attempting to use minutiae count as their quality benchmark (or attribute) and so are achieving different rankings to each other due to inconsistent determinations of the value of this attribute.

It may also be the case that there is variation between practitioners in the attribute used to compare quality. Some may be using minutiae count for the purpose, while others use clarity of marks, and others use type of pattern present, for example. Practitioners could also be using a more holistic quality determination during this comparison, differences in which could also lead to differences in the rank achieved (and, thus, the submission decision made).

Differences in target mark submission may be due to some practitioners using the immediate decision context upon which to base comparisons whereas others may have used exemplars from working memory (Stewart, 2009). As each of these outcomes is equally likely, this could explain some of the individual differences within the data. Equally, differences in target mark submission may also be due to the availability and type of marks in working memory. One practitioner using working memory as a source of exemplars, for example, may have just worked on a case involving many high quality marks (for example, cyanoacrylate developed marks on a plastic bag) whereas another may have been examining paper items with very poor quality, broken up developed ninhydrin marks, leading to a difference in the outcome of decision by sampling ranking due to differences in the quality of the comparison exemplars used.

There are slight differences between participants in the overall effect shown when considering the terminology of the inclusion/exclusion model (Bless & Schwarz, 2010). In the case of two practitioners an assimilation effect is shown to occur, whilst all other practitioners exhibit a contrast effect in their

mark submission decision making. This shows the importance of being aware that there are differences in the type of effect that background mark quality may have in operational submission decisions. From a psychological perspective it tends to be the case that an overall effect within the population (trend) is of interest, as it is in this application, but in this case it is also essential to acknowledge that the fingerprint submission task is an individual one. Each of these individual practitioner carries out this task and it is these individual decisions that govern the progression of evidence in a particular case. Therefore, if there is a minority who react differently in certain situations this needs to be acknowledged and considered. However, in these cases, it is important to ensure that this is not erroneously translated into an issue of error or blame (Frese & Keith, 2015), it is simply the case that there are individual differences in subconscious psychological effects.

#### 6.6.3. Objective 6.3: Assessing for the influence of demographic factors and the presence of order effects

Analysis of the effect of order of presentation of proportionate background mark quality and the effect of the position of target marks on the main effect of quality of background marks on target mark submission was possible due to the implementation of counter balancing measures in the experimental design. No significant difference was found between the main effect of background mark quality according to either the order of presentation of exhibits of differing background mark quality ( $p=0.632$ ), or the position of target marks ( $p=0.230$ ). This allows confidence in the main finding of a contrast effect in target mark submission without concern that this finding was influenced by either the order that the mock exhibits were presented or the position of the target marks within these exhibits.

The order of presentation of mock exhibits was unlikely to have had an effect on target mark submission, as a period of at least two weeks was left between participants carrying out each of the four experimental exercises (with the proportionate quality of background marks differing between each exercise). During these two week periods practitioners would have viewed many marks on live casework of varying qualities and quantities, and this should, and indeed was found to, be sufficient in avoiding any such order effects in the data.

The position of the four target marks within each exhibit was counter balanced using three different semi random positions of the target marks. Statistical analysis revealed that the position of these marks did not affect the overall effect of background mark quality ( $p=0.230$ ). This is an interesting finding as the decision by sampling (DbS) literature suggests that the decision may be influenced by the immediate context in which the target mark is presented (Stewart et al., 2006). Depending upon the definition of immediate context, this may suggest that the quality comparison undertaken in

relation to a target mark next to a poor quality background mark may be different to that undertaken in relation to the same target mark when placed next to a good quality mark within an exhibit of the same proportion of background mark quality. No significant difference in target mark submission according to the positioning of target marks may suggest that that sampling of the decision context occurs more widely than just the most immediate neighbouring marks to the target mark, or may suggest that the sample required for pairwise comparison scores to reach the necessary threshold for decision making is larger than the marks in the immediate vicinity of the target. A finding of no effect of the positioning of the target may also offer support for the inclusion/exclusion model and a mean determination of background quality being used during the comparative quality determination.

## 6.7. Conclusion

The results of this empirical study have demonstrated that the proportionate quality of fingerprints present on an exhibit at the point of 'marking up' may have an influence upon the submission decision made in relation to an ambiguous quality 'target' fingerprint on that exhibit. Overall, the presence of good quality background fingerprints reduced submission of ambiguous target marks, and the presence of poor quality background fingerprints increased submission of these target marks. This illustrates a trend towards a contrast effect during the mark submission process. This is an important finding as it may suggest a possible exaggeration of the findings of Chapter 3 and Earwaker et al. (2015) as, naturalistically, practitioners will submit a mark recovered based upon viewing it *in situ* upon an exhibit whereas an examiner will make a usability determination based upon viewing a mark in isolation.

# Chapter 7 Discussion

## 7.1. Overview of discussion

This discussion brings together the key findings of each chapter of this thesis and explains the interactions between these findings in relation to the fundamental aspects of the fingerprint submission decision process. This chapter first describes the problematic nature of, and inefficiencies within, fingerprint sufficiency decision making, in light of the findings of this thesis. A number of existing conceptual frameworks for discussing decision making problems are then described and a novel model for the consideration of the fingerprint submission decision is introduced. Key empirical findings and observations from throughout this thesis are then discussed in relation to each area of this novel model.

## 7.2. Inefficiencies within the fingerprint submission process

Initial research presented within Chapter 3 of this thesis found that there is the potential for inefficiencies in the fingerprint submission process from the fingerprint laboratory to the fingerprint bureau, when dealing with ambiguous (or borderline) fingerprints. This inefficiency is shown to occur in terms of discrepancies between practitioner submission decision outcomes and examiner fingerprint usability decision outcomes.

This is an important and, potentially, concerning finding for two key reasons. Firstly, it is important because it is highly undesirable to remove fingerprint evidence from the criminal justice system which has the potential to lead to an identification. A possible solution to this problem would be to implement a policy in the laboratory which states that practitioners should submit all pieces of ridge detail to the bureau (thus significantly lowering the threshold of practitioner mark submission). This, however, would be an inefficient process, leading to the submission of many marks of insufficient quality, and, without infinite (or indeed, in many cases, sufficient) resources available, this is not a financially viable option. Equally, this would also require a certain threshold in the detection of the presence of ridge detail which may be problematic (as discussed further later in this discussion in section 7.5.3). Instead, there is a requirement for the fingerprint submission process to deliver a filtering of fingerprint evidence that reflects the usability determinations of fingerprint examiners. Achieving this would result in an efficient process, delivering better value through an increase of evidence on the same budget.

Secondly, this finding is concerning because the production of forensic evidence needs to be a transparent process (Tully, 2015). If identifiable fingerprints are being discarded during evidence

recovery and submission then there needs to be transparency and justification in relation to this process. If identifiable fingerprints are being unintentionally discarded (as the findings of Chapter 3 may suggest) then this has potential implications for the reliability and admissibility of fingerprint evidence in a court of law.

### 7.3. Using decision making models to explore why the fingerprint submission process is not efficient

Empirical evidence from Chapter 3 of this thesis suggests that the fingerprint submission process may not be delivering the requirements of operational forensic science in efficiency and transparency, in relation to 'borderline' fingerprints. In order to better understand why this process is not successful, and how the findings of this thesis can be used to build a picture to aid explanation as to why it is not successful, the fingerprint submission process should be considered from a psychological perspective of judgement and decision making theory.

The fingerprint submission decision can be framed according to a number of different psychological models of decision-making. Three such models are introduced and discussed below.

#### 7.3.1. The principles of decision theory

Decision theory from a normative psychological approach, describes a decision in terms of the utility values and probabilities associated with it. The classic example of decision-making using decision theory describes the decision of whether or not to take an umbrella when going out. The model states that two pieces of key information are used when making this decision: the probability that it will rain, and the utility value placed upon having (or not having) the umbrella for each scenario (it rains or it does not rain). The decision maker takes into account how likely it is to rain and what value they place upon having the umbrella with them in each scenario and comes to a decision based upon the combination of these values.

Applying this model to the fingerprint submission decision would suggest the requirement for the equivalent two pieces of information (probability and utility) to be available to the decision maker. Probability can be considered in this case in terms of the probability that the mark will be useable to the examiner (as, if the mark is submitted it would be ideal that it is useable by the examiner in a similar way to if the umbrella is taken it is ideal that it would rain). Utility can be considered according to the value placed by the practitioner on each of the possible four decision outcomes:

1. The mark was submitted and it was useable
2. The mark was submitted but it was not useable

3. The mark was discarded but it would have been useable
4. The mark was discarded and it was not useable

Modelling the submission decision in this way requires consideration of the information required to establish probability and utility value which may be difficult to quantify given the high number of factors influencing the submission decision. For example, the utility value placed upon the mark being submitted but not being usable (a false positive submission) may differ according to the resources available in the case, the type of evidence upon which the mark was visualised, or submission policies existing within the laboratory.

### 7.3.2. The components of signal detection theory

The fingerprint submission decision can be considered based upon the principles of signal detection theory (outlined in 2.6.3). Signal detection theory discusses a subjective decision in relation to two key aspects; discrimination ability and decision threshold (Phillips et al., 2001). In terms of the laboratory practitioner submission decision the discrimination ability component can be considered as the discrimination (or the detection of) of features indicative of quality, for example determining the presence of a minutiae or 1<sup>st</sup> level detail in a developed mark. The concept of decision threshold in terms of the laboratory practitioner submission decision can be considered in two different ways. In terms of the meaning according to signal detection theory, threshold would imply the point at which a trace is discriminated from the background (i.e. the point at which a minutiae would be recognised as a minutiae). However, the term threshold is also of importance when considering making an overall submission decision. In this case threshold would refer to the point at which there is considered to be sufficient information, or quality indicators, present within the mark to make it suitable for submission. This threshold may be dependable upon operational requirements or resources available in the case.

### 7.3.3. Decision matching using Brunswik's Lens Model

Brunswik's Lens Model is used to describe the process that leads to a judgement on a problem that has a true solution or value, for example, determining the age or the profession of an individual of whom you have no prior knowledge, through the evaluation of the weight of a number of cues. Proportionate consideration of these weighted factors will lead to a judgement. Within the traditional Lens Model there are five key components that make up this judgement process; a criterion (for example, age of a stranger), cues (for example, hair colour), ecological validities (the reliability of the hair colour cue), cue utilisation (the weight placed upon hair colour information in the age decision), and a response or judgement (age determination) (Newell & Shanks, 2014). The Lens Model can be

applied to fingerprint submission decision making with a key adaptation; the judgement being made is not to determine a true solution or value existing within the world, it is, rather, to successfully determine (and mirror) the subjective judgement of another person (to determine and mirror the usability determination of a fingerprint examiner). This scenario means that the practitioner decision needs to utilise the appropriate cues and weights of these cues in order to make a submission decision appropriate to that of a fingerprint examiner. A novel adaptation of Brunswik's Lens model for the fingerprint submission decision is presented in Figure 4.14 (Chapter 4).

#### 7.4. Components and interactions within fingerprint submission decision making - a novel model

This novel model produced as part of this thesis combines key elements of the psychological models outlined, as is applicable to the fingerprint submission decision. The model also acknowledges the procedural and practical aspects of the mark submission process and combines this with simple terminology readily digestible by those without a psychology background. This combination produces a useful tool for consideration by operational fingerprint units, the wider forensic domain, the criminal justice system, and researchers when considering the key components of the fingerprint submission decision and how these relate to the wider environment in which the decision is made. The model is provided in Figure 7.1. The key components of the model are as follows:

##### Decision success

Decision success refers to the definition of a successful practitioner fingerprint submission decision. Decision success may be dependent upon policies and procedures of individual organisations, the resources available within a particular case, or the requirements of the fingerprint bureau in terms of use of the mark. As such there are many possible influences on decision success, and it may be dependent upon changeable thresholds.

##### Cues

Cues, as described within the model, refer to the detection of quality indicators within a fingerprint, similarly to discrimination ability within Signal Detection Theory. This encompasses aspects of visual perception and determining the presence or absence of a cue according to a threshold.

##### Thresholds

Thresholds in the model represent the point at which sufficiency is determined in a fingerprint in terms of a definition of decision success. An appropriate threshold for submission will vary according to decision success and, equally, desirable thresholds for submission may also influence the definition of decision success employed.



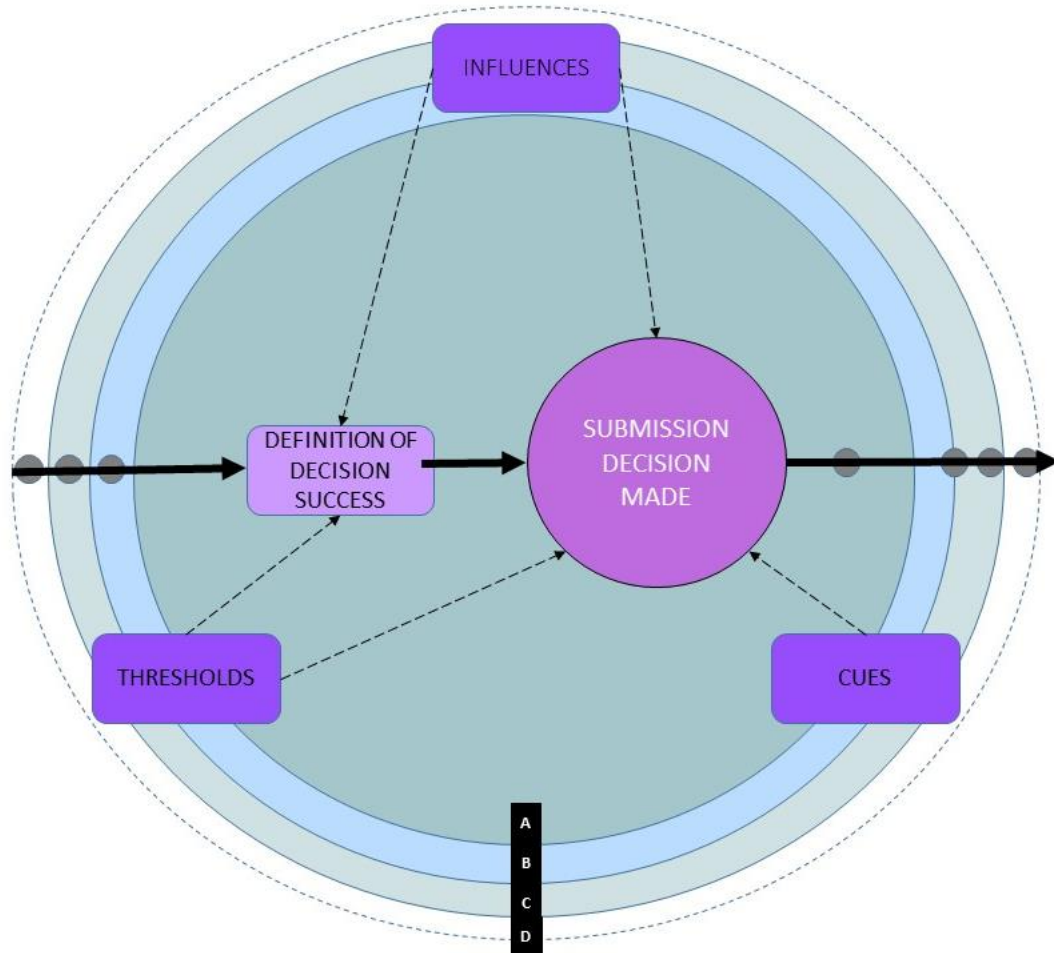
## Influences

Influences refers to any psychological factor which has an impact upon what makes a successful decision, and the actual decision made. Influences may be conscious or subconscious and may originate from a variety of sources from either the immediate ecology of the submission decision, or the wider ecologies of forensic science, the Criminal Justice System, or society.

## Ecology

In the model, ecology refers to the environment in which the submission decision is made. This may be the direct environment in which the decision is made in which aspects of the environment such as, the processes employed, or the relationship between the laboratory and the bureau may have an indirect impact upon the mark submission decision. The ecology of forensic science refers to aspects such as the impact of national best practice guidelines or legislation. The ecology of the criminal justice system refers to the impact of the requirement of the submission decision to be accepted and useful within the court system. The ecology of wider society refers to aspects of the wider views of society which may influence the fingerprint submission process, such as, for example, a belief that all ridge detail should be forwarded to the bureau, or the differing views of society in terms of the value of a conviction the case of different categories of crime. The horizontal arrow within the model represents the directionality of impact within these different ecologies throughout the decision process, with the grey dots emphasising impact within each ecology. For example, the requirements of the ecology of the forensic science may have an impact upon what is a successful submission decision if a probabilistic approach to the consideration of evidence is adopted. In turn, the outcome of the submission decision will impact in each ecology, for example, by adding to the body of forensic evidence in a case if submitted.

Figure 7.1 - Components and interactions within fingerprint submission decision making - a novel model



ECOLOGY of:	
<b>A</b>	Fingerprint Submission
<b>B</b>	Forensic Science
<b>C</b>	Criminal Justice System
<b>D</b>	Society

## 7.5. Discussion of the key findings of this thesis in terms of the novel fingerprint submission decision model

The empirical research contained within this thesis adds to the knowledge base within each section of the proposed model. The model is thus used as a framework for discussion of the key empirical findings of this thesis according to the five key components of decision success, cues, thresholds, influences, and ecology.

### 7.5.1. Decision Success

In order to investigate the efficiency and success rate of fingerprint submission, and to establish where this process may be failing, it is crucial to first establish what is meant by a successful submission decision, and how such a definition can be used as a benchmark to ensure successful decisions are made.

#### 7.5.1.1. The use of fingerprint examiners sufficiency decisions as an ecologically valid 'ground truth' measure of decision success

Throughout this thesis the usability decision of fingerprint examiners has been used as a benchmark for decision success with the assumption that the ultimate aim of mark submission is to submit all marks that are useable (or suitable) for comparison by an examiner and to discard marks which are not. However, it is acknowledged that variability in examiner fingerprint quality assessment and use of the ACE-V process has been shown in a number of studies (Dror et al., 2011, Fraser-Mackenzie et al., 2013) and was also demonstrated within empirical research within this thesis, both in overall aspects of usability determination during mark quality assessment in the preparation of experimental marks for use in Chapter 3 and in variation in minutiae count shown in Chapter 5. Ultimately it is highly challenging to assess the accuracy of fingerprint examiners quality assessments as there is no objective ground truth available upon which to base such a comparison. Therefore, a key assumption made within this thesis is that the quality assessment judgement of examiners was an appropriate measure for determining decision success in practitioner mark submission. This assumption was considered to be a justified approach given the lack of a truly objective measure, as fingerprint examiners are, essentially, the customer of the laboratory. This means that success in terms of the output of the laboratory could be defined as matching the required input of the fingerprint examiners. Whilst the variation and lack of an objective quality assessment (analysis) methodology amongst examiners is acknowledged, it was deemed that this 'ground truth' was ecologically valid. Variation in examiner sufficiency decision making will be present within operational mark submission and any examiner could be the customer of a practitioner submission decision.

In order to ensure that the 'ground truth' adopted in this thesis was representative of the population of examiners and not skewed towards a particular examiner or outlying viewpoint a consensus view of suitability for comparison was reached amongst five different examiners in relation to each experimental mark about which usability was required as an analytical benchmark for empirical study (experimental marks used in Chapter 3, 4, and 5, and 'target marks' used in Chapter 6).

The concept of decision success in terms of examiner suitability can be neatly illustrated through the novel application of Brunswik's Lens Model to this decision outcome matching scenario, in which the components of the practitioner's submission decision should mirror the components of an examiners usability decision in the case of a successful decision. This relationship is illustrated in Figure 4.2 (Chapter 4).

#### 7.5.1.2. Variation in decision success

The definition of decision success in mark submission may vary according to a number of different factors.

Firstly, as previously discussed, the examiner benchmark for sufficiency is subjective with variation in general aspects of sufficiency decision making (Dror et al., 2011, Fraser-Mackenzie et al., 2013) and also there are differences in the objective measures which may influence sufficiency, for example variation in minutiae count as found in Chapter 5 of this thesis). This may mean that success varies according to the examiner or bureau receiving the submitted mark.

Success can also vary as a result of a number of different procedural and technological requirements associated with the fingerprint comparison process. For example, if AFIS is to be used to search for possible database matches then this may alter the threshold of mark quality required. Decision success would then, in this situation, be the submission of a mark of high enough quality to be searched on AFIS. On the other hand, a direct manual comparison of a crime scene mark against the tenprint card of a suspect is likely to be less objective in its requirements (and again more subjective from the view point of an examiner's quality assessment). The requirement for either an AFIS search or direct tenprint comparison will be influenced by a variety of investigative factors in the case such as the type of crime or the presence of suspect or elimination prints. There is, therefore, a discussion to be had about the requirement for case related information to be passed to laboratory staff. Depending upon the set up and systems of working between the laboratory and the bureau it may be the case that laboratory staff are made aware of the existence of suspects or elimination prints in an investigation via the paperwork or electronic records submitted with the evidence requiring fingermark enhancement. In the current environment in which an increasing importance is placed upon the recognition of the possibility of cognitive biases within the forensic sciences (Dror et al., 2006, Dror et

al., 2011) and a number of proposed solutions to such biases include the removal of task irrelevant information (Dror, 2009, Wells et al., 2013), it is important to be aware of such situations in which information which may be considered to be irrelevant in relation to the immediate task (mark enhancement and submission in this case) may actually be of importance in gauging decision success. It could be considered, in this situation, that the use of a case manager (as has been proposed (Dror, 2016)) may be a sensible approach for ensuring important information is available to the practitioner so that they know what decision success in fingerprint submission looks like, but that this information is presented in a way which is less likely to lead to subconscious biases. For example, a practitioner may be told by a case manager that the marks submitted in a case need to be suitable for search on AFIS, but do not necessarily need to be told the crime type of the case, unless this is an additional factor which would procedurally alter mark submission thresholds (further discussed in 7.5.3).

There is also currently discussion around a shift to the increased use of a probabilistic approach to reaching conclusions and the reporting of these conclusions by fingerprint examiners (Neumann et al, 2015). Whilst this approach is yet to be fully implemented within operational case work, the use of probabilities in the fingerprint comparison process has the wider potential to impact upon the definition of decision success in relation to the mark submission of laboratory practitioners. This is because the adoption of a probabilistic approach signals a move away from the concept of a 'match' being the primary information delivered by the fingerprint examiner. Instead an examiner may be able to apply a statistical approach to communicating the value of smaller quantities of ridge detail meaning that an absolute threshold does not need to be crossed. If an examiner is using such a probabilistic approach to communicate the results of a comparison, then the definition of a successful submission decision would then be measured by the mark being deemed of sufficient quantity and quality for this novel assessment, presumably reducing the threshold for mark submission. It could be considered that a move to an entirely probabilistic approach would mean that any piece of ridge detail may be able to provide information. This could, arguably, remove the necessity for the fingerprint evidence filtering process carried out by laboratory practitioners. This is an important area for future consideration should a probabilistic approach to reporting fingerprint comparison conclusions be more widely adopted.

In addition to technological and methodological aspects of the fingerprint comparison process, there may be financial reasons for variation in the definition of mark submission decision success. If more resources (in terms of both staff and materials) are available within the budget of an investigation then this may mean that the law enforcement organisation decides that they are prepared to lower the threshold for mark submission in order to guarantee that evidence is not missed. This could take

the form of a mandate directing practitioners to submit all enhanced fingerprint detail in a case, leading to a changed definition of decision success.

Such differences in resource availability may be targeted according to the perceived seriousness or impact of the crime that has taken place. For example, a murder or terrorism case may have a higher budget for forensic evidence and allow for submission of a higher number (and lower quality) of fingerprints. Arguably, in this situation, the definition of decision success remains the same – success is still submitting useable fingerprints, but the tolerance for error in decision making has increased with the reduction of the submission threshold (to be discussed further in 7.5.3. in relation to thresholds), meaning that success from the view point of the role of the practitioner in this process has changed. It could also be argued that differences in the definition of submission decision success according to crime type or circumstances can be said to be dependent upon the utility value placed upon the outcomes of the investigation by society. If society places a higher value on apprehending an offender in a murder case than on apprehending a burglar, then this will influence prioritisation of forensic resources. The lay community of potential jurors also have the potential to influence what a successful submission decision looks like. The jury can be seen to be the ultimate customer of the fingerprint recovery process as it is they (alongside the judge) who will utilise the fingerprint evidence to draw conclusions about innocence and guilt in the case. In order to draw sound conclusions, the judge needs to be sure of its admissibility of forensic evidence in court in order to provide the jury with reliable evidence. The expectations of these stakeholders and the admissibility standards of the court could govern what a successful submission decision is. For example, a lay audience or a legal gatekeeper may have the expectation that all ridge detail enhanced during a fingerprint examination is forwarded to a fingerprint examiner, particularly in the case of a serious crime. The discarding of any form of fingerprint evidence at this stage may be considered to be erroneous decision making. As such it is crucial that decision success is defined by laboratories and that the parameters that effect this definition are set out. Furthermore, it is essential that this definition of decision success is based upon sound empirical study. An empirically established definition of submission decision success enables transparency in the mark submission process, and for a measure of reliability to be established.

Whilst the empirical research contained within this thesis has centred around a definition of success in mark submission which reflects the usability determination of examiners, it is recognised that there are additional factors which have the potential to influence decision success. The aim of practitioner submission decisions matching practitioner usability decisions, however, seems to be a sound starting point for the exploration of this process.

### 7.5.2. Cues

The term cue in this context is used to describe the indicators of quality (or, more specifically, the indicators of 'submissibility') present in a fingerprint. The idea of the use of cues in decision-making can be described according to the principles of Signal Detection Theory. Cues (or signals) first need to be detected, then a threshold applied to indicate that the cues establish the presence of a target (in this case the presence of a fingerprint suitable for submission). Given that empirical work included in Chapter 3 of this thesis illustrated discrepancies between the submission decisions of laboratory practitioners and the usability determinations of fingerprint examiners (in the case of borderline marks) it was crucial to aim to better understand how practitioners were attempting to make successful fingerprint submission decisions, in order to further establish where these discrepancies lie. An important first step in this process was an examination of decision cues, indicative of mark quality.

#### 7.5.2.1. Qualitative insight into the decision cues used by practitioners

The findings of qualitative empirical research exploring the decision rationale of laboratory practitioners was presented in Chapter 4 of this thesis. These findings provide an insight into the importance of certain cues, or quality indicators, when making a fingerprint submission decision. Some cues were found to be more commonly utilised by practitioners. The clarity of a mark was found to be important, as were rationale related to the characteristics present within the mark. Indeed, there were many references to a numerical count of characteristics (minutiae), with practitioners commonly stating that there were 'enough' or 'sufficient' characteristics present to submit a fingerprint. Some practitioners went a step further and stated the value of characteristics upon which their submission decision was based. In responses provided to the experimental study in Chapter 5 of this thesis, in which practitioners were asked to state the number of minutiae present in a series of fingerprints, two of the participating practitioners only provided a minutiae count up to a certain threshold, one implying, and the other stating in text, that this was because once they had reached this minutiae count they would submit the mark, and so never needed to count beyond this threshold in case work. The findings of the study in Chapter 4, combined with these incidental findings of the study presented in Chapter 5 illustrate the apparent importance of minutiae as cues to indicate mark suitability for submission. In addition, anecdotal evidence gleaned during discussions with practitioners revealed a commonly held assumption that, if a practitioner could see a certain number of minutiae, they would submit a mark because they believed that an examiner would be able to see more minutiae. Whilst these findings are important in the consideration of thresholds (discussed further in 7.5.3) they are also important indicators of the importance of the use of minutiae number as a decision cue.

Indeed, the use of minutiae count as a more objective cue to indicate mark quality would seem, perhaps, a logical and natural approach to mark submission that may reduce some of the ambiguity surrounding the quality of a fingerprint. However, when asked about policy in relation to mark submission no laboratories questioned stated that they had a policy based a numerical value of minutiae. Given the apparent discrepancy between policy and practice it seemed important to investigate the potential usefulness of minutiae as a decision cue through further empirical study.

#### 7.5.2.2. Quantitative investigation of the potential value of minutiae as a decision cue

The empirical findings presented in Chapter 5 of this thesis found that, when asked to count the number of minutiae present in marks of a range of qualities, there was considerable variation within both the practitioner and examiner groups. Indeed, the largest range of responses in relation to the same mark within the practitioner group was 34 and within the examiner group was 37.

There was also no statistically significant difference between the minutiae counts of examiners and practitioners made in relation to the same sets of fingerprints. The range of minutiae counts given by the practitioners and the examiners is arguably, however, one of the most important pieces of information gleaned from the study. This is because, on a decision by decision basis, differences in average populations and standard deviations are arguably unimportant as the real world process involves one lab practitioner making one decision for the benefit of one examiner. That practitioner could sit anywhere across the range of minutiae counts within the practitioner group, as could the examiner across the examiner group. There is no guarantee which examiner-practitioner combination will be struck. This means that the two may have similar or very different approaches to counting minutiae. There would certainly be no benefit in this situation to applying a rule of 'if the practitioner can see X minutiae, the examiner will be able to see X +Y' as the range and sign of Y is likely to differ between cases.

The findings of Chapter 5 suggest that there may be little value in the use of minutiae count as a submission decision cue as a minutiae count does not deliver the objective methodology that may have been expected, rather the process of determining the presence of minutiae, even in good quality fingerprints, is subjective for the human decision maker. Determining a minutiae count at this stage in the fingerprint evidence recovery process could, perhaps, be a task for a computer programme in the future (similar to the 'marking up' capabilities of the AFIS system). However, this would require further empirical research to establish a correlation between examiner usability determinations and objective minutiae count provided by such technology. If no such correlation exists then minutiae count may not a useful cue upon which to base submission decisions.



### 7.5.2.3. The potential use of other cues to improve the efficiency of the fingermark submission decision

The process of fingermark quality assessment can be considered to be a subjective process using a range of different cues or quality indicators and this is problematic to a wider extent than just within operational casework involving mark submission from the laboratory. A particular area in which a more consistent approach to mark quality assessment between individuals would be beneficial is in the grading of fingermarks for research purposes. The more reproducible and objective a methodology for grading the fingermarks developed during research projects in which the aim is to assess the comparative effectiveness of different development techniques, the more easily studies can be replicated and results from one study can be directly compared to those of another study carried out by different individuals or across institutions. Given the potential advantages of a consistent approach to this 'grading for research' task there have been a number of attempts at developing research methodology and grading schemes for this purpose (Sears et al., 2012, Fieldhouse, 2009, Humphreys et al., 2008). A recent study reports an attempt to train researchers in the use of a fingerprint grading scheme and introduced a proficiency test for completion following training (Fieldhouse & Gwinnett, 2016). Student participants were generally found not to be able to demonstrate a satisfactory level of proficiency in the use of the grading scheme (Fieldhouse & Gwinnett, *ibid*). This highlights the highly subjective nature of the fingerprint grading task, even when a more objective methodology is provided and when training in the use of this methodology is also given. Laboratory practitioners who have never received any official training in fingermark quality assessment and do not have a documented or objective methodology would perhaps be expected to employ a subjective methodology resulting in individual differences in performance. Fingerprint grading for research purposes could be argued to sit outside of the requirements of operational process as the ultimate purpose of this grading is a comparison of the quality of the product of chemical enhancement, not an assessment of the usability or suitability for comparison of the mark. As such the grading system used by Fieldhouse and Gwinnett (2016) features cues of mark quality including the area of developed mark present and the level of contrast between the mark and the background surface. It could, however, also be potentially beneficial to consider the cross over between fingermark grading for research and fingermark quality assessment for mark submission decisions. Ultimately, the product of the application of fingermark visualisation methods is ridge detail for use in the fingermark recovery and comparison process. Therefore, it seems sensible that the performance of a development technique is judged against a set of cues which are applicable to the mark submission and comparison processes. This would enable research projects to be collaborative between both multiple academic institutions and also between academic institutions and operational fingerprint units, adding value to the performance conclusions

that could be reached in relation to new techniques in terms of operational usability. To achieve this aim a crucial starting point is a further consideration of the most important cues both within the mark comparison process and also in terms of comparing the cues of successful mark development. The production of an objective grading scheme which takes into account both of these key factors, in particular any cues that overlap between the two aims would seem to be a potentially useful extension of research into fingerprint grading.

### 7.5.3. Thresholds

Throughout this thesis the term threshold is used to describe, as per signal detection theory (Phillips et al., 2001), the point (threshold) at which the cues or quality indicators in a mark are deemed to be sufficient (in quantity/quality or both) to enable a decision to be made in relation to the 'submitability' of a mark.

#### 7.5.3.1. Overall findings in relation to practitioner submission thresholds

It was initially hypothesised within Chapter 3 of this thesis that any discrepancies between practitioner fingerprint submission decisions and examiner mark sufficiency decisions may have been due to an inappropriate overall threshold being adopted by practitioners (for example consistently either submitting at too high a threshold (making majority of decisions 'errors' by discarding usable marks) or consistently submitting at too low a threshold (making the majority of decision 'errors' by submitting unusable marks). Such an effect would have demonstrated in the data in a majority of either false positive or false negative decision errors within the practitioner group. This, however, was not found to be the case. Rather, there was an equal proportion of false positive and false negative errors present within the data. This would suggest that the concept of thresholds in relation to mark submission may be, in fact more complex than the use of an overall, broad threshold of general quality. It may instead be the case that different thresholds are being adopted in different cases, be this in relation to different marks, or between different practitioners. This inconsistency in practitioner submission thresholds highlights the problematic and subjective nature of applying a threshold to this suitability decision.

#### 7.5.3.2. Improving the objectivity and specificity of the submission threshold

One of the key objectives of Chapter 5 of this thesis was to ascertain if a comparative threshold of minutiae counts could be utilised to improve the efficiency of the mark submission process. Such an approach would have the potential to allow a comparative quantification of the submission threshold in terms of the number of minutiae present within the mark. For example, were it to be shown through empirical study that examiners are able to determine the presence of X more minutiae in a fingerprint

than practitioners then this would enable the implementation of a numerical minutiae count threshold for practitioners which would ensure that examiners were likely to be forwarded marks with sufficient minutiae for comparison (at a pre-agreed value). Such an approach would allow simple changes in threshold in order to meet a changing definition of decision success. For example, if decision success was the submission of marks which were of sufficient quality for uploading on the AFIS system then this would mean increasing the threshold of practitioner submission to the point at which the practitioner- examiner minutiae count ratio would mean that examiners would be likely to view sufficient minutiae present in the mark for uploading to AFIS.

The hypothesis that such methodology may have the potential to improve efficiency of the mark submission process was based upon data collected during the empirical study and upon anecdotal evidence obtained throughout the research process. There was a consensus amongst the laboratory practitioners involved in the research that a minutiae count threshold was an important aspect of their mark submission process. This was demonstrated in the qualitative research in Chapter 3 of this thesis which outlined minutiae count as a reason for a submission decision, with a threshold value of minutiae being specified by some practitioners. Additionally, during the study detailed in Chapter 5 two practitioners and one examiner responded to the experimental task of counting the minutiae present in a mark by stating when their minutiae count had exceeded their personal threshold for submission or use. Anecdotally some practitioners were prepared to provide further qualification of this approach, stating a belief that if they could see X minutiae an examiner could see X + Y minutiae in the same mark. These findings provided evidence for the belief that a minutiae count threshold was an important factor used within the mark submission process. The empirical study detailed within Chapter 5 of this thesis, therefore, looked to establish an empirically sound basis for the use of a minutiae count based submission threshold.

It is important to remember that fingerprint examiners in the UK do not use a numerical standard for the purposes of fingermark comparison since the abolishment of the 16 point standard in 2001 (Mackenzie et al., 2011). Whilst the formalisation of the use of a numerical threshold of minutiae may seem to be a lean towards such a sole reliance on a numerical standard, it is important to recognise the distinction between the use of a numerical minutiae count for mark submission as opposed to for mark comparison purposes. The use of a numerical threshold for mark submission does not necessarily imply a reliance upon a minutiae count during fingermark comparison. Fingerprint examiners may be using any cues for quality assessment and comparison, the practitioner minutiae count value would simply allow practitioners an objective methodology for submission determination at an objective and easily manipulated threshold according to the requirements of the force, casework, and examiner.

However, of course, if minutiae count has no bearing on, or correlation with, examiner mark usability at all, then such an approach would only lead to further inefficiencies in the mark submission process.

The findings of Chapter 5 of this thesis indicate, however, that the practical issues of applying a minutiae threshold as a more objective methodology of mark submission may be a purely academic discussion. The empirical study detailed in this chapter found considerable individual differences in the minutiae count achieved for the same mark within both practitioner and examiner groups. No relationship in minutiae count was identified that would support the anecdotal hypothesis that if a practitioner can see X minutiae an examiner can see X + Y. Such considerable variation in minutiae count, even in clear, non-ambiguous marks, and a lack of a relationship between examiner and practitioner minutiae counts would suggest that attempting to use a formal minutiae count threshold as a submission threshold would only led to increased potential error in the submission process.

It is important to bear in mind that this thesis has focussed upon the threshold related to one particular quality cue (minutiae count). It may well be the case that practitioners use a number of different thresholds when making mark submission decisions, or take a holistic approach through the use of an overall quality threshold. Further empirical study is needed to explore both naturalistic and optimum thresholds.

Such further research is particularly vital given that there is currently no documented threshold, criteria, or mechanism for mark submission. The lack of an agreed threshold or criteria means that there is no objective criteria upon which to judge success in submission decision making, or to provide feedback against. Rather, the only means of assessment in this skill is according to the subjective determination of the fingerprint examiner. Within the current culture of accreditation, proficiency testing, and transparency, such an approach feels outdated and problematic. The lack of a documented mechanism and threshold would suggest that fingerprint laboratory practitioners are carrying out a subjective task relying upon their expertise without the training process or feedback allow the development of such expertise. It would seem that either more objective and measurable thresholds need to be put in place, with the organisation determining what makes a successful submission decision, or there needs to be a shift in the way in which the role of the laboratory practitioner is viewed, moving away from a role as a technician towards acknowledgement as, and the necessary support provided to become, an expert (as per the view of an expert witness).

The use of a disclosed and empirically valid threshold for submission has considerable scope as a mechanism for ensuring that submission decisions are successful in a landscape of changing decision success. This is because having a set threshold means having a threshold that can be lowered or raised to deal with changing operational requirements, for example different resources available within an

investigation, or the different end requirements of a fingerprint (e.g. AFIS search vs manual comparison).

In addition to the future consideration of useable objective thresholds it is also interesting to consider the potential role of decision theory in better understanding and manipulating the naturalistic decision thresholds of laboratory practitioners. A key component of decision making from a decision theory perspective is the utility value placed on a particular outcome. Further research could be valuable in assessing the effect of manipulating the utility values placed on a submission decision on the threshold of submission.

#### 7.5.4. Influences upon decision making

There are many conscious and subconscious influences that may affect each component of the decision processes, from the determination of decision success, to the cues which are representative of mark quality, and the thresholds required to indicate mark sufficiency. Such influences may be beneficial or detrimental to the process of decision making and the outcomes of this process. Influences may be inherent in the decision making process itself or may be external, stemming from the ecology of the decision, or the wider ecology of forensic science.

##### 7.5.4.1. Subconscious influences

Subconscious influences upon decision making within forensic science and fingerprinting have been widely discussed and empirically investigated within recent years. Research in this area has been pioneered by several scholars (Found, 2014). Research into the effects of cognitive bias, whilst initially controversial, has been carried out within a number of forensic science disciplines including DNA (Dror and Hampikian, 2011), fingerprinting (Dror et al., 2006), blood pattern analysis (Laber et al., 2014), forensic odontology (Osborne et al., 2014), forensic anthropology (Nakhaeizadeh et al., 2014), handwriting analysis (Stoel et al., 2014), ballistics (Kerstholt et al., 2010), and shoeprint analysis (Kerstholt et al., 2007). Indeed 'bias' has become a buzz word within forensic research, conferences, and meetings, emerging as its own category as 'Cognitive Forensics', and becoming a motivational factor for organizational change (Dror, 2016).

Within fingerprinting such research into cognitive bias has focussed upon the more commonly viewed interpretative aspects of the fingerprint evidence process, commonly focussing upon the study of cognitive biases within the fingerprint comparison (ACE-V) process (Dror et al., 2006). Whilst fingerprint comparison is, indeed, a highly interpretative element of the fingerprint evidence process, it is important to acknowledge that there are interpretations and subjective and ambiguous decisions made throughout the progression of fingerprint evidence, from the consideration of potential

fingermark evidence at the crime scene to the presentation of fingermark evidence in court. These subjective decisions (including fingermark submission, both from the crime scene and the laboratory) have the potential to be affected by the same cognitive biases as the more traditionally viewed interpretative aspects of the fingerprint evidential process as these are challenging decisions which are often made under time pressure, within an emotional context, and without an objective methodology (Dror et al., 2006). As such it is important to investigate the presence and impact of cognitive biases throughout the fingerprint evidence process so as to better understand where cognitive biases may be occurring and to appreciate the impact that this may have upon the efficiency and accuracy of the evidence recovery process.

Some, however, have argued that a disproportionate focus of research into subjective decision-making is preventing the forensic domain from focusing on increasing the objectivity with which forensic evidence can be interpreted, through a better understanding of the forensic traces themselves (Champod, 2014) through, for example, empirical studies aiming at understanding the persistence and transferability of such traces (Scott et al., 2014, Morgan et al., 2014). Others have responded by arguing that there is not, in fact, an overrepresentation of research into cognitive bias within forensic science, and that research establishing the existence and prevention of cognitive bias and the more objective use of trace evidence need not be mutually exclusive (Risinger et al, 2014). It would, indeed, seem to be sensible that empirical study looks to fill the knowledge gaps in relation to both the physical science of forensic traces and the processes used to analyse these, and that it also seeks to better understand the interactions between the human scientist and these traces and processes. It is only by understanding all of these varied aspects of the forensic process that we can fully justify the weight and meaning of our inferences in case work.

#### 7.5.4.2. The effect of crime type on mark submission

Subconscious influences were explored in relation to the impact of crime type on fingermark submission decisions during the experimental work detailed within Chapter 3 of this thesis. Crime type was deemed to be a potential subconscious (as opposed to a conscious) influence upon decision making as, within the participating laboratories, there was no policy for marking up exhibits differently according to the type of crime scene from which the exhibit had been recovered. This lack of a differential policy would suggest that the crime type was not, therefore, a conscious consideration when making a submission decision. The empirical study detailed in Chapter 3 investigated the effect of crime type on the submission rates and threshold of fingerprint laboratory practitioners. Differences in mark submission were examined according to whether practitioners had been told that the marks originated from a volume crime of 'theft from vehicle' or from a serious crime of 'murder'. Differences in mark submission according to the crime type given were only found in one of the

participating laboratories. The data shows that practitioners working within a laboratory that routinely dealt only with exhibits from serious crimes had reduced their submission threshold and had submitted less marks when dealing with exhibits from the mock volume crime during the experimental task. This suggests that crime type may only have influenced the decision making of practitioners when they were carrying out the more uncertain task of mark submission in relation to a crime type that they did not normally deal with. This may suggest that crime type will have a greater effect the more ambiguous and challenging a decision is, perhaps meaning that crime context would have a greater effect on fingerprint quality determination in non-experts who would rely more heavily on such information in the absence of expertise or experience. It would be interesting to carry out further work in this area to assess the differences in the effect of crime type information on fingerprint quality assessment, this may be of particular importance when considering the application of researcher grading schemes to operational casework and differences in the use of contextual information between researchers and practitioners.

Crime type, as previously discussed, can be viewed as a valid organisational motivating factor for a change in the definition of submission decision success and, consequently, a change in submission threshold due to the increased value (both societally and therefore financially) placed upon solving serious crimes. However, in this present research carried out within three UK police forces it was found that none of the participating laboratories had a policy which indicated a difference in approach to marking up according to crime type.

However, practitioners did suggest that crime type was crucial information required to make a submission decision. During the completion of the empirical study presented with Chapter 6 of this thesis participants were asked to carry out a naturalistic marking up task using the mock experimental exhibits provided. The instructions provided during this task deliberately made no reference to crime type; they simply stated that practitioners should 'mark up the exhibit as [they] would during casework'. Upon receipt of the experimental material practitioners, however, requested further crime type information in relation to the mock exhibits, stating that they needed this information in order to be able to carry out the marking up task. When clarification was sought from the laboratory manager it was disclosed that there was indeed no policy or requirement that stated that there were differences in the way that marks should be submitted according to the category of the crime. Nevertheless, practitioners felt that this was essential information for their decision making process. Given the lack of policy, and the lack of an overall effect of crime type on mark submission found in the study detailed in Chapter 3 within a police force that dealt with both types of crime, this would suggest that neither are practitioners expected to adjust their submission processes or thresholds, nor do they in practice. This suggests the importance of further empirical study to establish the part played

by crime type in the decision process, perhaps through the application of a decision theoretical approach to the submission decision.

#### 7.5.4.3. The effect of background mark quality

Research included within Chapter 6 of this thesis provides empirical evidence that the quality of background marks on an exhibit can affect the decision to submit or discard ambiguous marks on that same exhibit.

This is an important finding for two key reasons. The first is that this finding identifies background mark quality to be a subconscious influence in the decision making process. Prior to this research the quality of 'background' marks present upon an exhibit has not been known to, or been considered to, influence fingerprint submission decision making in relation to a 'target' mark. It could be that the use of the quality of other marks on an exhibit to make a holistic decision about the submission of a particular ambiguous mark could be considered to be a sensible approach. A consideration of the quality of marks present on an exhibit as a whole allows a practitioner to attribute limited resources available in a case towards the marks that they believe have the best chance of being usable for identification purposes by an examiner. However, if this approach is to be taken then this needs to be a conscious approach documented in procedure, such as a 'take the 20 best' procedure. Such an approach also needs to be objective. The occurrence of a contrast effect, as discussed within Chapter 6, suggests a non-deliberate prioritisation of marks for submission according to the quality of the background marks present. Further investigation here is important to better understand the cognitive processes underlying this effect. A key question to ask is whether this is purely a subconscious cognitive effect (or heuristic) or whether there are elements of operational culture (such as a high requirement for efficiency) which are driving individuals to make this prioritisation. Consideration of this contrast effect by policy makers could lead to a determination that it is likely to be operationally beneficial and efficient to be adopting an approach which bases submission upon the quality of other marks with an exhibit, whether this be a subconscious influence or a conscious consideration, but, arguably, consideration needs to be given to the transparency of such a process to ensure admissibility in court and prevent legal challenge.

The second reason that a context effect in mark submission is important is because of the operational reality that submission decision success is measured as a consistency between the outcome of a practitioner fingerprint submission decision and an examiner usability determination. As practitioners tend to make submission decisions based upon a mark surrounded by other marks visualised upon an exhibit, and examiners tend to make a usability determination based upon the mark in isolation, a finding of an effect of background mark quality on the quality determination of borderline fingerprints



may provide some explanation for discrepancies in the outcome of the decision making between the two stakeholder groups. It may be that procedural change leading to practitioners making submission decisions in relation to marks in isolation may help to increase decision success, although the increase in the efficiency achieved may not outweigh the additional time and resources required to adopt this procedural change (for example, the requirement to isolate each mark for decision making, possibly through initial image capture of each individual mark prior to decision making).

#### 7.5.4.4. Increasing our understanding of influences

Empirical studies within this thesis have explored the potential of crime context and background mark quality to influence mark submission decision making. It is important to bear in mind that there are many other potential influences on this decision making process, each requiring empirical investigation in order to gain a more comprehensive picture of their impact. A starting point for the investigation of such influences should be a comprehensive map of the submission decision and the external influences present. A thorough review of the psychological literature in relation to these types of influence should then be undertaken in order to set the agenda for future applied empirical studies. This thesis has paved the way for such future work, not only in fingerprint submission but also at a wider level within fingerprinting and forensic science. It is clear that further research is required to fully appreciate the effect, and better understand the impact, of cognitive influences on decision making within forensic science, particularly within the 'hidden' decisions and interpretations made throughout the forensic process. It would seem that future research in this area needs to be grounded in both operational procedure and the science of judgement and decision making so as to produce findings which are both empirically robust and also operationally relevant and beneficial to the reliability, transparency and efficiency of forensic processes.

#### 7.5.5. Ecology

In the present model the term ecology is used to describe the environment in which the fingerprint submission decision sits. This can be the environment of the submission decision itself (the ecology of fingerprint submission), the wider ecology of forensic science, or the broad ecology of the criminal justice system. The impact of the findings of this thesis are discussed in terms of each of these ecologies.

##### 7.5.5.1. The ecology of fingerprint submission

The observations and empirical findings contained within this thesis have a direct impact upon the ecology (or process) of fingerprint submission. Empirical study within this thesis has highlighted the potential for identifiable marks to be prematurely removed from the process of fingerprint recovery.

This, along with observations of current practice, suggests the need for a feedback mechanism which can detect both false positive submissions and false negative submissions, which is currently lacking within the standard approach to workflow. Such a feedback mechanism should replicate an agreed definition of decision success. For example, if decision success is measured as a submission decision mirroring an examiner's usability decision, then a feedback mechanism which detects 'errors' as per the view point of an examiner needs to be in place. Feedback is well recognised as being essential for the development of expertise (Ericsson & Lehmann, 1996) and may well be a simple method of improving the efficiency of the current mark submission process.

Empirical findings of this thesis have evidenced the challenging, subjective, and interpretative nature of submission decision making and fingermark quality assessment. This is in contrast to the prior commonly held view of this process as an objective task that does not require expertise in relation to fingerprints and fingerprint quality assessment. This view of fingerprint visualisation as a technical specialism as opposed to an interpretative expertise is evident in the limited coverage that fingermark quality assessment is given in either training (Lagden, 2014), home office guidance (Bowman, 1998), proficiency testing programmes (Stow, 2014), or ISO accreditation requirements (Hall, 2014). The present findings suggest the potential of training, guidance, best practice recommendations, proficiency testing and accreditation in improving the efficiency and performance of laboratory fingerprint quality assessment. However, such inclusions need to be based upon a solid empirical backing and an interdisciplinary approach to truly understanding the decision process, along with an agreement upon a societal and operationally acceptable definition of decision success. As such the content of this thesis provides a starting point for future endeavour within the ecology of laboratory fingermark submission.

Indeed, the empirical findings of this thesis have had an impact upon the ecology of fingermark submission to date. One UK metropolitan police service has introduced new methods of working to enable the dip sampling of exhibits so as to allow the evaluation of discarded fingermarks in terms of the judgement of a fingerprint examiner. This is as a direct result of empirical findings of inefficiencies in this area as set out in Chapter 3 of this thesis. In addition, fingermark submission decision making is now considered as an important process in relation to ensuring quality within the fingermark laboratory, and is set out as an area for future inclusion within the forthcoming Forensic Regulator's Codes of Conduct for Fingerprint Laboratories.

The ecology of forensic science

The empirical findings of this thesis have implications wider than the fingermark visualisation laboratory. Scene of Crime Officers can be considered to be carrying out a very similar decision task

to that of laboratory practitioners when determining which areas of ridge detail that they have visualised at a crime scene that they will lift or image and submit to the fingerprint bureau. Should the results of empirical research within Chapter 3 of this thesis be applied to the scene of crime officer mark submission process then this may suggest a considerable proportion of marks that could lead to an identification being lost through not being recovered from the crime scene. Arguably, such issues within this process could have larger implications than those found to occur within the laboratory. This is due, in part, to the current lack of a mechanism for picking up on false negative submission errors (marks left at the crime scene), and the challenges associated with the implementation of such a mechanism. Crime scene examiners will often visit a volume crime scene alone, and, once they have carried out their investigation, the property will be returned to the occupant who will typically clean any surface upon which fingermarks have been developed. As such it is more so the case at the crime scene that the decisions made at the point of evidence recovery will be the only opportunity for the recovery of this evidence. Furthermore, applying psychological decision theory to the consideration of this problem may suggest that crime scene examiners would place a higher utility value on not forwarding poor quality marks to the bureau, than on leaving good quality marks at the scene, as it is only upon the former criteria that their performance can be judged, given the lack of a mechanism for the detection of false negative decisions. Further empirical study looking at the efficiency and mechanism of mark (and wider evidence) recovery at the crime scene would be a valuable extension of this thesis.

The findings of this thesis have also highlighted the importance of acknowledging that interpretative and subjective decisions are made throughout the forensic process, and are not limited to the tasks commonly considered to be interpretative in nature (for example, fingerprint comparison, or the interpretation of a DNA profile). The conclusions of this thesis highlight the potential for improved efficiencies in other domains of forensic science through the study of decision making as applied to particular human decisions throughout the entirety of each forensic process, particularly focussing on these early interpretations made during evidence recovery which have largely been overlooked to date.

#### 7.5.5.2. The Ecology of the Criminal Justice System

Appreciating the breadth and potential impact of human decision making throughout the entire forensic process is crucial to ensure that forensic science delivers a high quality service with integrity and transparency. Without a full and open understanding of where errors due to human decision making may be occurring within this process full integrity and transparency cannot be achieved. More specifically, within fingerprint submission, the direct implications of the empirical findings of this thesis are of fundamental importance as they may suggest that a shifting role of the fingerprint

enhancement officer in court, from providing a technical account of the processes carried out in the laboratory to an expert providing opinion evidence which includes a justification of marks submitted and discarded in the case, is required in order to maintain the integrity and transparency of the specialism.

It is important to have sound understanding of how the decision making process translates into fingerprint identifications in court when thinking about novel ways of presenting and discussing the weight of such evidence. For example, a move to a more probabilistic approach to the communication of findings within fingerprint evidence, would require a knowledge of the probability that fingerprints had been developed either in the laboratory or at the crime scene that could have been used to identify an offender but had not been recovered. This would become important information to take into consideration when drawing inferences with regards to fingerprint identifications presented in the case. A key defence line of questioning may well be ‘what is the chance that the fingerprint of another suspect was also developed at the crime scene and was not recovered?’ In order to be able to answer this question, which, given the findings reported within this thesis, may well be necessary as it would be quixotic for all pieces of developed ridge detail to be recovered, further empirical study is required to ascertain a best practice approach for which a reliable error rate can be calculated (and provided to the court) and against which practitioners can demonstrate proficiency and competence.

When discussing the wider legal implications of the findings of the present research it is important to highlight a key barrier to the progression of new ways of working in this area and one of the key challenges within the communication of such research. This key barrier to progress is the challenges associated with identifying and discussing error within the forensic and, particularly, human identification fields. There are two fundamentally problematic aspects to the discussion of error within the forensic sciences. Firstly, there is the problematic nature of a culture of error prevention (as opposed to error management) which stems from the idea that all error can (and indeed should) be prevented. Such an approach has the potential to lead to a culture of individual blame, which prevents open and honest discussion of errors, which could lead to the management of such errors, and a reduction in their impact.

Empirical findings of errors within forensic processes (such as those described within this thesis) can have the potential to lead to discussions around the blame of individual incompetency for such errors. Often, and certainly in the case of the research described in this thesis, this misses the point of the findings. During the course of the experimental work it was, indeed, suggested that the results of the research could be used as performance indicators. Whilst there may be some scope for such an adaptation to research activities for the future use in proficiency testing, this inappropriately diverts

attention to an individual as opposed to an overall performance level. The empirical data gleaned within this thesis repeatedly demonstrated individual differences in task performance. However, these individual differences did not show some practitioners to be performing badly, it was rather the case that practitioners were performing differently, making erroneous submission decisions in different cases. This finding of individual differences highlights the subjective and challenging nature of the mark submission task, and highlights the differences inherent in the human decision maker. It is these differences that need to be acknowledged, and accepted in order to in turn accept that all human decision makers will make errors. It is only through accepting this and moving to a culture in which these errors (and the empirical research which explores them) can be openly discussed without fear of recrimination or stigmatisation that organisations such as fingerprint laboratories can learn to manage these errors and improve performance.

The idea of error management hits a further obstacle when it comes to expert evidence and the legal system. Indeed, the second fundamentally problematic aspect of discussing error within forensic processes is the requirement of forensic science to provide reliable scientific evidence in a court of law. Openly discussing empirical research (for example through peer reviewed articles) which exposes errors occurring in forensic processes has the potential to override the ability of the scientist to do just this as it casts doubt upon the integrity of the process or the scientist in the public domain. However, without empirical study forensic science cannot establish where improvements can be made which can lead to greater efficiency of processes and an ultimate increase in integrity and transparency.

# Chapter 8 Conclusion

## 8.1. Overview of key findings

This thesis set out to establish the efficiency of the fingermark submission process across UK fingerprint laboratories, to examine the fingerprint laboratory practitioner fingermark submission decision in terms of consideration of the decision cues used by the practitioners, and how these relate to the decision cues used by examiners, and to establish the effect of the contextual factor of background mark quality on the outcome of fingermark submission decisions.

Empirical studies carried out and reported within this thesis have highlighted potential inefficiencies in the fingermark submission process which could lead to the loss of identifiable fingermarks and resources wasted on the submission of unusable marks. These findings were replicated across two UK fingerprint departments. Submission decision success (in relation to fingerprint examiner usability determinations) was found to vary according to the fingermark being assessed, and inter-practitioner variation was found in relation to the same fingermarks. Practitioners tended to be overconfident in their decision making in relation to their decision accuracy. It was found that the type of crime committed can (but does not always) effect mark submission.

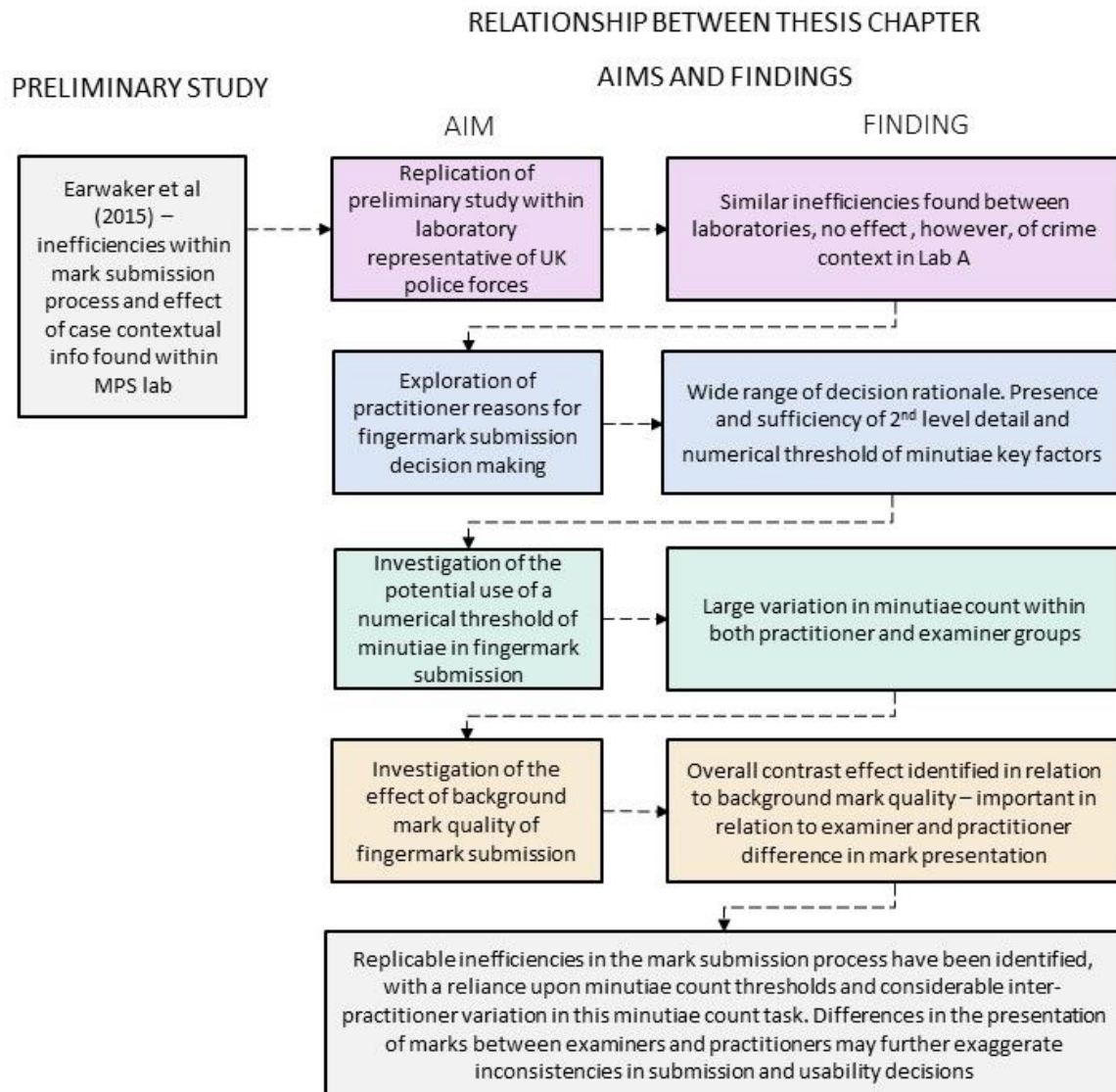
Laboratory practitioners stated 2<sup>nd</sup> level detail to be the most common reason for making a mark submission decision, heavily relying on a numerical minutiae count threshold, whilst also using the presence or absence of a pattern type within a mark and the clarity of a mark to inform decision making. Examiners, whilst also utilising 2<sup>nd</sup> level detail, were found to focus more upon the continuity of both ridge flow and minutiae count. Individual differences in decision rationale were found within both the practitioner and examiner groups.

A significant level of variation in minutiae counts was found within both practitioner and examiner groups. In addition, no statistically significant difference was found between the minutiae counts of practitioners and examiners according to experimental fingermark.

The proportionate background quality of fingermarks present upon an exhibit affected the level of submission of ambiguous fingermarks present upon the same exhibit. A contrast effect was observed; a large proportion of good quality fingermarks present upon an exhibit led to lower submission of ambiguous, but identifiable fingermarks, whereas a large proportion of poor quality marks present led to higher submission of ambiguous marks.

The overall aims and key findings of this thesis are outlined below in Figure 8.1.

Figure 8.1 - Relationship between thesis aims and key findings



The full findings of the thesis according to each research objective for each experimental chapter are outlined below.

### 8.1.1. Key findings of Chapter 3

Objective 3.1: Efficiency of fingerprint sufficiency decision-making within Laboratory A

The main findings identified were that:

- 33.65% of the practitioner submission decisions made during this study were erroneous (in relation to examiner usability determinations about ambiguous fingerprints). 34.29% of these were false negative decisions; 60 identifiable fingerprints were discarded during this study.

- Practitioners made more errors in relation to fingermarks of insufficient quality than in relation to sufficient quality marks.
- There was a variation in decision outcomes according to experimental fingermark.
- Erroneous decisions were made by all practitioners. Whilst there was some variation in decision outcomes, the errors made were divided between practitioners and were not due to a small minority of practitioners.

Objective 3.2: Effect of contextual information on sufficiency decision-making within Laboratory A

The main findings identified were that:

- There was no statistically significant difference in decision outcomes according to the case context in which the fingermarks were presented. However, there were inconsistencies in decision making in relation to the same marks when the crime context was changed; a higher number of inconsistent decisions were to submit a mark when it was presented in a serious context, and to discard the same mark when it was presented in a volume context, as opposed to the other way around.
- There was variation in the decision consistency and type of decision inconsistency observed according to the experimental fingermark.
- All practitioners were mostly consistent and correct in their decision making although there were some individual differences in decision inconsistency.

Objective 3.3 – The relationship between self-reported confidence and decision accuracy within Laboratory A

The main findings identified were that:

- Practitioners reported themselves to be less confident when discarding fingermarks, making an erroneous decision, and making a false negative decision.
- The majority of practitioners were overconfident in their sufficiency decision making.
- Practitioners were overconfident in their submission decisions made in relation to all but six of the experimental fingermarks. Practitioners were more confident in relation to these six marks, and these were the six with the highest decision accuracy.

Objective 3.4 – Comparison between mark submission within Laboratory A and the Metropolitan Police Serious Crime Laboratory

Objective 3.4.1: Comparison of the efficiency of fingermark sufficiency decision-making

The main findings identified were that:



- The percentage of 'correct' decisions (according to examiner usability determination) was the same for both forces. The two laboratories therefore also made the same proportion of erroneous decisions (33%)
- The erroneous decision of the two forces have a different make up. Laboratory A made a higher percentage of false positive decisions (22%) and a lower percentage of false negative decisions (12%) than the MPS laboratory, which showed an even divide of error types at 17% each.
- Decision making outcomes per fingerprint showed similarities between the two forces
- There are similar individual differences between practitioners in both forces

Objective 3.4.2: Comparison of the effect of contextual information on sufficiency decision making

The main findings identified were that:

- There is a significant difference in the effect of crime context on decision outcomes between the two laboratories
- In the case of the Laboratory A there was no significant difference in decision outcomes for the two crime contexts, but there was a significant difference found in decision outcomes according to context within the MPS laboratory
- The MPS laboratory was less consistent in its decision making than the Laboratory A (although this was not found to be a statistically significant difference)
- Where there was inconsistency in decision making the MPS laboratory submitted marks in a serious case and discarded the same mark in relation to a volume case more than the Laboratory A, and discarded marks in a serious case whilst submitting the same mark in a volume case less than the Laboratory A

Objective 3.4.3: Comparison of the relationship between self-reported confidence and decision accuracy

The main findings identified were that:

- Similarities in mean calibration were noted for practitioners from both laboratories
- The majority of practitioners from both laboratories were found to be overconfident in their decision making
- There was no correlation found between confidence and accuracy according to practitioner for either laboratory

- There were similarities in calibration according to fingerprint demonstrated by the two laboratories
- In the case of both laboratories there was a weak visual correlation between accuracy and confidence according to experimental fingerprint
- Both laboratories demonstrated good correlation between confidence and accuracy in relation to the fingerprints upon which they been most accurate, and the identity of these marks was common to both laboratories
- Both laboratories were over confident in their decision making in relation to the majority of fingerprints

#### 8.1.2. Key findings of Chapter 4

Objective 4.1. To explore the rationale behind laboratory practitioner fingerprint submission decision making within Laboratory A.

Objective 4.1.1. A qualitative assessment of practitioner self-reported submission decision rationale within Laboratory A

The main findings identified were that:

- A broad range of rationale were reported by practitioners when deciding whether to submit or discard an experimental mark
- Most commonly reported categories of rationale were associated with 2<sup>nd</sup> level detail and clarity of the mark.
- Within the category of 2<sup>nd</sup> level detail, factors which indicated that a count or threshold of minutiae was important made up a total of 78% of responses

Objective 4.1.2. An assessment of inter-practitioner variation in fingerprint submission decision rationale with Laboratory A

The main findings identified were that:

- The make-up of decision categories across individual practitioners was largely representative of the overall make-up of decision categories
- There were some individual differences in the type of decision rationale reported
- There were individual differences in the number of rationale reported
- The 'no response' decision factor was more common amongst the responses of a minority of participants

Objective 4.1.3. An assessment of practitioner fingerprint submission decision rationale according to experimental fingerprint

The main finding identified was that:

- There are similarities in the proportion of self-reported decision categories according to experimental mark

Objective 4.2. An investigation of inter-laboratory consistency in practitioner decision rationale

The main findings identified were that:

- There were overall similarities in the decision categories observed across the two metropolitan UK police force laboratories
- There was a difference in the relative proportion of the 'clarity' and '1st level detail' decision categories between the two laboratories, with the MPS laboratory reporting more decision factors within the '1<sup>st</sup> level detail' category, than the 'clarity' category, and the reverse being true of Laboratory A
- There were overall similarities in the make-up of decision categories given for both categories of crime by both laboratories
- In the case of the MPS Laboratory there was a decrease in the proportion of the '2<sup>nd</sup> level detail' and the 'clarity' categories and an increase in the proportion of 'no response' responses from the serious to the volume crime category.

Objective 4.3. Exploring the relationship between practitioner submission decision rationale and fingerprint examiner usability rationale

Objective 4.3.1. An assessment of fingerprint examiner self-reported usability decision making rationale

The main findings identified were that:

- Fingerprint examiners reported a range of rationale for fingerprint usability decision making
- Decision factors reported were most commonly related to the use of 2<sup>nd</sup> level detail, whilst clarity and movement in the mark was also a key category of rationale

Objective 4.3.2. A comparison of fingerprint practitioner submission decision rationale and fingerprint examiner usability decision rationale

The main findings identified were that:

- There were overall similarities in the categories of decision rationale stated by examiners and practitioners
- There was, however, a difference between the proportion of responses in relation to 1<sup>st</sup> level detail and the clarity of a fingerprint, with examiners relying more heavily on factors relating to mark clarity and practitioners relying more heavily on factors associated with the presence of pattern or 1<sup>st</sup> level detail

### 8.1.3. Key findings of Chapter 5

Objective 5.1: An examination of inter-practitioner variability in fingerprint minutiae count

The main findings identified were that:

- There is considerable variation in practitioner minutiae count in relation to the same experimental fingerprints
- The range of minutiae counts is often high with the most commonly occurring range of counts (in relation to five fingerprints) being 14
- There is considerable variation in total minutiae counts in relation to all marks stated by each practitioner. Total minutiae counts vary from 110 to 604 minutiae

Objective 5.2: A comparison of variability between practitioners and examiners

The main findings identified were that:

- Overall there was similar variability within the practitioner and examiner groups
- On a per fingerprint basis practitioners demonstrated a larger standard deviation in minutiae count than examiners in relation to 38 of the 60 experimental fingerprints

Objective 5.3: A comparison of practitioner and examiner minutiae counts

The main findings identified were that:

- Overall there was no statistically significant difference between the minutiae counts of practitioners and examiners
- Mean minutiae count of examiners was, however, higher in relation to 40 of the fingerprints whilst higher mean minutiae counts were observed within the practitioner group in relation to the remaining 20 marks

Objective 5.4 – The effect of fingerprint quality on minutiae count and variability

The main findings identified were that:

- The number of minutiae observed by both the examiner and practitioner groups increased as the quality ranking of the marks increased
- Examiners observed more minutiae than practitioners in relation to borderline and good quality fingermarks whilst practitioners observed more minutiae in relation to poor quality marks
- When per fingermark data was arranged according to mean minutiae count both practitioners and examiners illustrated greater variation as mark quality (in terms of minutiae count) increased

Objective 5.5 – Inter-laboratory consistency in practitioner variability and minutiae count

The main findings identified were that:

- There was no significant difference in the minutiae counts of practitioners according to the laboratory in which they worked
- Laboratory B exhibited a slightly higher variation in minutiae count than Laboratory A

8.1.4. Key findings of Chapter 6

Objective 6.1. - An investigation of the main effects of background mark quality on target mark submission

The main findings identified were that:

- There was a decrease in the mean number of target marks submitted as the proportionate quality of the background marks present increased
- There was a significant difference between the number of target marks submitted when exhibits which contained an extreme good quality background context (100% good quality marks) and each of the other proportions of background mark quality
- There was no difference between the effect of the proportionate background mark quality on target mark submission between the two laboratories

Objective 6.2. An assessment of individual differences in mark context effects

The main findings identified were that:

- There are considerable individual differences in the number of target marks submitted across the experimental conditions
- There is some individual difference in the effect of context mark quality on borderline mark submission

Objective 6.3. Assessing for the influence of demographic factors and the presence of order effects

The main findings identified were that:

- No exhibit order or target mark positioning effects were found
- There was no significant difference in context effect according to gender, or years of experience within the laboratory

## 8.2. Implications of the empirical findings of this thesis within fingerprint laboratory fingerprint submission decision making

The empirical findings of Chapter 3 of this thesis have suggested that fingerprint submission decision making and the progression of fingerprint evidence from the laboratory to the bureau could be a more efficient process. This is a crucial finding as efficiency within fingerprinting (and, indeed, other areas of forensic science and criminal investigation) is crucial for delivering a best value service. Contrary, perhaps, to the majority view of the general public and the requirements of the current College of Policing training programme for laboratory practitioners (Lagden, 2014), the recovery of all items of developed ridge detail is not a financially viable, or realistic proposition. The findings of this thesis, however, suggest that increased efficiency can be achieved without the requirement to submit all visualised ridge detail. This is due to the observation of an equal proportion of false negative and false positive erroneous decisions by practitioners. Consequently, making better decisions could potentially increase the number of usable fingerprints submitted to the bureau, without requiring the straightforward lowering of a submission threshold which would be likely to result in the submission of a higher number of fingerprints overall. This finding is important as it suggests the potential financial benefits of practitioner engagement with such research. Equally, the finding of the potential for identifiable fingerprints to remain unrecovered from an item of evidence is important as, by accepting that this can occur, a more open discussion around a definition of decision success and an approach to manage, as opposed to only prevent, such errors can take place.

Empirical study within Chapter 4 of this thesis has established that there is a lack of an agreed mechanism or protocol for making a fingerprint submission decision and that there is variation in the mechanisms being used by practitioners. This finding is important as it suggests that regulatory bodies and government organisations should take care in the provision of quality standards that mandate that an organisation should have, and follow, an internal policy in relation to the selection of fingerprint ridge detail for submission (ENSFI, 2015, and Forensic Science Regulator, 2016a). As there is currently an absence of such a policy (established within Chapter 3 of this thesis), and a lack of clarity has been demonstrated in the way in which such decisions should be made to ensure that they result

in an efficient mark submission process (established through empirical study presented within Chapter 4 of this thesis), it would seem that further empirical study with the 'buy-in' from fingerprint units is required in order to be able to provide an empirical basis for fingermark submission decision policy, which could then be mandated.

Whilst there would be significant value in further empirical research to comprehensively explore and further understand current, and optimum, mark submission decision making mechanisms, the novel empirical findings of this thesis can be used to begin to inform laboratory practice. Many of the key findings could be used by fingerprint departments when considering best practice in fingermark submission. Firstly, the findings of a practitioner overreliance in the use of 2<sup>nd</sup> level detail, and the use of an unofficial numerical threshold in minutiae count could be discussed. Explaining to practitioners the importance of sequential minutiae and ridge flow from an examiner perspective may be a starting point to improving decision success. Secondly, implementing a regular feedback mechanism in terms of examiner assessment of discarded marks may be a relatively simply implementable process, which could address a number of findings of this thesis. Such a process could provide simple outcome feedback or, more comprehensive, cognitive feedback to practitioners, developing their level of expertise, encourage two-way dialogue in relation to key quality cues within mark quality assessment, and, perhaps, lead to improved calibration of practitioner confidence. A feedback mechanism of this kind may also help to improve the working relationship between laboratories and bureaux and help to foster a culture of error management. Thirdly, research presented with Chapter 3 and 6 of this thesis has established that the current process of fingermark submission may be affected by information which is present during decision making, both in terms of an indication of the crime type of the case (Chapter 3, section 3.4.3), and the presence of 'background' fingermarks upon an exhibit (Chapter 6). The necessity of such information within the mark submission process should be considered by laboratories. It may be that the disadvantages of removing such information (similar to the lights out processes suggested by Kassin et al. (2013)) and adapting the working practices necessary to do so, may outweigh the potential gains in increased efficiency in the mark submission process. Therefore, operational trials and further empirical study should first be carried out in order to assess the costs and benefits of such an approach.

### 8.3. Implications of the findings of this thesis within the wider process of fingerprint evidence recovery

The findings of the present thesis have highlighted a potential gap in the training received by fingerprint laboratory practitioners in terms of how to determine which fingermark ridge detail to capture and submit to the fingerprint bureau. It would seem that an acknowledgment of the

operational reality that decision success does not involve the submission of all visualised ridge detail by key training stakeholders is crucial in order to deliver fit for purpose training in a relation to fingerprint submission. Care should be taken, however, to ensure that any changes to practitioner training are based upon appropriately design empirical findings and agreed best practice and definitions of decision success. It may be that a national approach needs to be taken, or that it is agreed that there are differences in requirements at a local level (for example differences in workflow and casework types) which mean that training needs to be tailored to individual forces or collaborations. A starting point for improvements in fingerprint submission decision making training would, perhaps, involve multi stakeholder discussion between the newly established national laboratories working group, the College of Policing and key training partners, the Forensic Science Regulator's fingerprint quality standards working group, and academic stakeholders.

A key consideration is the implication of all findings of the present thesis to the wider process of fingerprint evidence recovery. It is important to acknowledge that laboratory practitioners are not the only stakeholders making fingerprint submission decisions on a regular basis as part of their job role and without the benefits of specific training in fingerprint quality assessment. Scenes of crime officers regularly decide whether or not to submit powdered and lifted fingerprints from the crime scene. This, arguably, is a very similar decision to that of laboratory practitioner fingerprint submission with, however, a number of notable potential differences. There could be considered to be the potential for a greater effect of heuristics and cognitive biases within the scene of crime officer submission decision due to the potential for increased ambiguity, time pressure, and exposure to emotional situations (Dror et al, 2006). Additionally, the crime scene is, by nature, a less controlled environment with less opportunities for feedback in mark quality assessment, either in terms of the judgment of a fellow crime scene investigator or a crime scene manager, or in terms of the usability determination of a fingerprint examiner. This may suggest that there could be an even greater disparity between the submission decisions of scenes of crime officers and the usability determinations of fingerprint examiners particularly in the case of ambiguous fingerprints, given the importance of feedback in the development of expertise (Ericsson & Lehmann, 1996). Further empirical study to apply the principles of the methodological approach of Chapter 3 of this thesis (and Earwaker et al., 2015) to fingerprint recovery by crime scene investigators at the crime scene is an important further extension of this thesis.



#### 8.4. Implications of the findings of this thesis for fingerprint and wider forensic science research

The findings of Chapter 5 of this thesis that demonstrate significant inter-practitioner variation in, what would seem to be, the objective task of fingermark minutiae counting, support the findings of Fieldhouse and Gwinnet (2016) that the quality assessment of fingermarks, even through the application of an apparently objective grading method is a challenging and subjective process. It would, however, seem that there may be advantages in pursuing the use of objective methodologies for fingermark quality assessment and empirically exploring different approaches to achieving consistency and decision success through the use of objective criteria. Achieving proficiency in mark quality assessment through the use of objective criteria could be mutually beneficial to both operational fingermark submission and also to cross institutional fingerprint research (as suggested by Fieldhouse and Gwinnet, *ibid*). In addition, the approaches of fingermark grading for research and operational fingermark quality assessment could be combined in order to facilitate a comprehensive approach to interdisciplinary and inter-institutional research that allows the direct application of empirical findings to operational casework.

Indeed, experimental work carried out throughout the present thesis is an example of successful academia-practitioner partnerships which have combined the observation of a problematic process within operational casework, the application of empirical research through a multidisciplinary academic approach, and consideration and dissemination of findings which may lead to real world improvements in process and performance. The potential impact of the findings of this thesis, in terms of potential increased efficiency in mark submission through the consideration of discrepancies between practitioners and examiners in the mark submission process, highlights the potential strength of such research collaborations and illustrates that sound empirical research based upon psychological principles and research which looks directly at operational processes need not be mutually exclusive.

Indeed, an important aspect of the present research is that it has highlighted that, in ambiguous cases, the wrong fingermarks may be being submitted and discarded by the laboratory. Knowledge of this inefficiency (gained in Chapter 3) and some of the factors that may be contributing to it (such as a reliance of minutiae count thresholds, large inter-practitioner variation in minutiae count, and contrast effects in relation to background mark quality) demonstrate the potential to add value to the fingerprint evidence recovery process through further consideration of the decision process. This is a key consideration for research within fingerprinting, and the wider forensic science domain, at a time financial resources are stretched and budgets for research and development are limited. Having

reached a number of empirical findings in relation to psychological effects within the fingerprint submission decision it is crucial that these findings are peer reviewed within the academic arena, but also that the findings are disseminated to an audience of operational stakeholders. As such, the findings of each chapter of this thesis will be published in academic journals, and will also be disseminated to management and practitioner stakeholders within the participating laboratories in the form of a series of workshops (see List of Publications). Ensuring that the end product of the research undertaken reaches an operational audience at the appropriate levels is key to ensuring that the benefits of taking part in the collaborative project are realised. Benefit realisation is a key aspect of successful project management (APM, 2012) and is concerned with ensuring that the end product of the project is fully utilised. In the case of the all the research presented within this thesis benefit realisation could be achieved by working with operational stakeholders in order to implement changes in policy such as the introduction of more comprehensive feedback mechanisms, or making submission decisions in isolation (away from the background quality of marks present). However, continued empirical study is required to be able to fully realise the benefits of this work to date and to make recommendations for best practice in fingerprint submission decision making that may lead to a better value fingerprint recovery process. It is crucial that collaborative research within forensic science is project managed to include benefit realisation in order to successfully contribute to empirical knowledge and also to successfully inform practice, providing good value for both academic and practitioner stakeholders.

Publishing the findings of academia-practitioner collaborative research within the scientific community is essential to allow peer review and also to enable continuation of research themes across academic groups. However, there are a number of challenges inherent in publishing academic research that involves forensic practitioner participants. These challenges include issues around the disclosure of findings which may suggest errors or inefficiencies within forensic processes, and the disclosure of information identifying a police force or organisation. Publishing such findings and information is problematic as it has the potential to expose issues within a process to the general public and legal profession which could lead to media attention or challenges in court. It is crucial, therefore, that publishing performance related empirical findings is carried out sensitively, in order to balance the benefits of transparency and peer review with the potential costs in terms of negative publicity or the potential challenge of evidence in court. In a practical sense this can be achieved through a collaborative approach between academic and practitioner stakeholders to publication writing and final approval, or through the use of disclaimers clarifying the difference between casework and experimental findings. In addition, participating practitioner organisations can choose to remain anonymous in publications. A close working relationship between researchers and

practitioner stakeholders throughout a project can be beneficial to gain trust in relation to publication, and a pre agreement that data collected during a project can be published can help to ensure that peer review is possible.

#### 8.5. Implications of the findings of this thesis for wider forensic science and the legal system

The implications of all the findings of the present thesis should be considered in terms of the wider process of forensic evidence recovery from crime scene to court. This thesis has highlighted that there are subjective and interpretative decisions being made earlier in the chain of evidence recovery than is, perhaps, commonly considered. This highlights that interpretation is not confined to the traditionally interpretative aspects of the evidential process (fingerprint comparison in the case of fingerprint recovery) but that it occurs throughout the process from the recovery of evidence to the crime scene to the presentation of evidence in court.

Indeed, this thesis has highlighted that the role of the fingermark enhancement laboratory practitioner is wider than selecting and carrying out chemical treatments and that the role requires elements of subjective judgment. In addition, the empirical findings of this thesis have highlighted that practitioners do not routinely recover all items of developed ridge detail from an exhibit, and have a variety of reasons for submitting or discarding a fingermark. These findings have the potential to have an impact upon the role of the laboratory practitioner in court, as it may be that the questioning in relation to fingermark recovery moves to focus upon the more interpretative aspects of mark recovery, leading to more 'expert witness' style questioning for the practitioner. It may be that it is beneficial to provide training to practitioners in relation to such potential cross examination.

There is also scope for the findings of this thesis to inform evidence recovery in other areas of forensic practice due to their potentially transferable nature. For example, the finding that the quality of background fingermarks can influence the quality assessment (in terms of submission) of ambiguous fingermarks could have transferable properties to the gender determination decision made by a forensic anthropologist in relation to an ambiguous bone. If the bone being assessed is surrounded by other bones with clear female features, it may be that a contrast effect occurs making the 'target' bone appear less clearly female in its traits. Care should be taken, however, that such findings are not directly transferred to other decisions within the forensic science process, rather that they are used to signpost decisions and processes in relation to which further targeted empirical study may be of benefit, given the complex differences inherent in each forensic process.

## 8.6. Implications for other domains

The present thesis also has potential applications within a number of domains external to forensic science, both as a direct result of the findings of the empirical research carried out and also in relation to the approach taken to build positive interdisciplinary research links within a potentially controversial area of research.

The novel application of psychological theory to the fingerprint laboratory practitioner submission decision presented within this thesis has the potential to contribute to the wider psychological literature surrounding expert decision making. The research findings of Chapters 3 – 6 allow an increased understanding of decision making within an area of professional forensic specialism which would be traditionally challenging for psychologists to gain access to. These studies can also be considered to provide a starting point for the further application of psychological research methods to the fingerprint submission decision and decisions throughout the wider forensic process both in terms of a starting point of a knowledge base, and in terms of the further exploitation of the established partnerships with key UK fingerprinting stakeholders. Key examples of further psychological research that could be mutually beneficial to increasing the knowledge base within decision making psychology and improving the efficiency of fingerprint submission processes are the application of signal detection methodology to the fingerprint submission process, the further exploration of psychological approaches to understanding metacognition, and the consideration of decision thresholds from a perspective of psychological decision theory. This thesis has provided a starting point for further research within these, and other, specialist areas.

The empirical findings throughout this thesis could also be of value when considering decisions made within domains other than fingerprinting and forensic science. A key example of a similar decision process to laboratory practitioner fingerprint submission is the patient triage decision made by triage nurses in accident and emergency departments within the medical domain (Cooke and Jinks, 1999). In this decision process the triage nurse must, essentially, determine which patients to refer to a doctor urgently and which less urgent cases can wait to be seen. Similarly to the practitioner fingerprint submission decision, whilst the triage nurse is a medical expert, he or she does not have the same expertise as the doctor who will eventually make a diagnosis. Consideration of this process relation to the novel application of Brunswik's lens model provided in Figure 4.2 frames the decision in terms of the importance of matching the decision cues between the nurse and the doctor, as it is the role of the nurse to essentially match the 'severity of illness' decision of the doctor. Similarly to the laboratory practitioner submission decision, resource constraints mean that that the optimum position of the nurse's triage decision is not to forward all patients for immediate treatment, rather

to only submit those who require it. At a basic level there can be considered to be four potential outcomes of a binary triage decision; to correctly refer a patient in need of urgent treatment, to successfully hold back a patient who is not in need of urgent care, to incorrectly refer a patient who is not in need of urgent care, and to incorrectly hold back a patient who is in need of urgent care. The Manchester Triage System widely used in UK Accident and Emergency Departments (Cooke & Jinks, 1999), however, highlights a key difference between the medical triage system and the laboratory practitioner mark submission decision. Whilst both the laboratory practitioner and the triage nurse are, essentially, attempting to replicate the view of an expert (the fingerprint examiner, and the doctor, respectively) the laboratory practitioner is making a binary decision (to submit or to discard a fingerprint), whereas the triage nurse must assign the patient to one of five categories of waiting time according to the severity of the case. These categories range from 'immediate' (a wait of zero minutes) to non-urgent (a maximum wait of 240 minutes). However, there are similarities in the tendency for the assigning of patients to the correct triage category to be overlooked in terms of performance statistics and validation which focus, instead, upon the proportion of patients in each category and the time taken to treat them (Goodacre et al., 1999, Cooke & Jinks, *ibid*), similarly to a focus upon fingerprint examiner decision making within fingerprint recovery. Given the similarities between the nature of the two processes it may be that the empirical studies contained within this thesis may be of benefit to inform future empirical study within the domain of medical triage decision making, for example considering the application of a study looking to compare triage nurse and doctor use of The Manchester Triage System, investigating the relationship between confidence and accuracy within triage decisions, or assessing the effect of the order of the presentation of patients with varying degrees of severity of symptoms on the classification of triage decisions made.

## 8.7. Summary of research impact

This thesis has taken a collaborative and interdisciplinary approach to investigate a specialist area of fingerprint recovery in which previously there was a lack of published research. In doing so, the value of research that applies scientific knowledge to operational practice through empirical studies with the potential of adding value to forensic processes has been highlighted. This research has identified challenges faced in terms of carrying out efficient fingerprint submission decision making and has illustrated the complexity of this process, previously considered to be straightforward in nature. These findings have highlighted the need to approach fingerprint laboratory mark submission from a scientific and psychological perspective in order to better understand how this process is carried out by human decision makers, and how it can be improved to increase the efficiency of fingerprint recovery. A need to explore fingerprint submission decision making at a wider level within fingerprint evidence recovery (particularly within scene of crime officer evidence collection) has been highlighted,

as has the importance of considering the role of interpretative decision making in evidence recovery across domains of forensic science throughout the forensic process from crime scene to court.

It is intended that this research will form the basis of continued interdisciplinary empirical study to translate the findings presented in Chapters 3 – 6 of this thesis into empirically tested recommendations that will inform policy in relation to training, and best practice within fingermark submission decision making. It is hoped that the findings and discussion of this thesis will go some way to foster a culture of error management and transparency in relation to the human aspects of fingermark recovery, and will set a precedent for successful collaborative research in this area between laboratory practitioners, fingerprint examiners, and academic stakeholders.

## List of publications

### Current publications:

Earwaker, H., Morgan, R. M., Harris, A. J. L., & Hall, L. J. (2015). Fingermark submission decision-making within a UK fingerprint laboratory: Do experts get the marks that they need? *Science and Justice*, 55 (4), 239-247.

Earwaker, H., Charlton, D., & Bleay, S. (2015). Fingerprinting – the UK Landscape: Processes, Stakeholders and Interactions. *Forensic Science KTN*.

### Publications for submission in 2016:

Earwaker, H., Morgan, R., Harris, A., Hall, L. (2016). A comparison of the fingermarks sufficiency process across UK police forces. *Science and Justice*. Submission Oct/Nov 2016.

Earwaker, H., Nakhaeizadeh, S., Smit, N., Morgan, R. (2016). The importance of applying judgment and decision making theory to the forensic evidence recovery process from crime scene to court (editorial paper). *Science and Justice*. Submission Oct/Nov 2016.

Earwaker, H., Morgan, R., Harris, A. (2016). An investigation of the submission decision making rationale (metacognition) of laboratory practitioners, establishing key factors in fingermark signal detection. *Science and Justice*. Submission Nov 2016.

Earwaker, H., Harris, A., Morgan, R. (2016). An assessment of the effect of the quality of background fingermarks on borderline quality mark submission. *Journal of Applied Research in Memory and Cognition*. Submission Nov/Dec 2016.

Earwaker, H., Morgan, R., Harris, A. (2016). A comparison of minutiae recognition ability and consistency between fingermark laboratory practitioners and fingerprint examiners. *Forensic Science International*. Submission Dec 2016.

## References

- Abraham, J., Champod, C., Lennard, C., & Roux, C. (2013). Modern statistical models for forensic fingerprint examinations: A critical review. *Forensic Science International*. 232(1–3), 131-150.
- Abraham, J., Champod, C., Lennard, C. & Roux, C. (2013a). Spatial analysis of corresponding fingerprint features from match and close non-match populations. *Forensic Science International*. 230, 87 – 98.
- Adams, J. K. (1957). A confidence scale defined in terms of expected percentages. *American Journal of Psychology*. 70, 432 – 436.
- Adams, J. K. & Adams, P. A. (1961). Realism of confidence judgements. *Psychological Review*, 68, 33-45.
- Ames, D. R., Kammrath, L. K., Suppes, A. & Bolger, N. (2010). Not so fast: The (not- quite-complete) dissociation between accuracy and confidence in thin-slice impressions. *Personality and Social Psychology Bulletin*. 36, 264–277.
- Association of Project Management (APM). (2012). APM Body of Knowledge. 6<sup>th</sup> Edition. Ibis House. Buckinghamshire.
- Armelius, B. A. & Armelius, K. (1976). Confidence and performance in probabilistic inference tasks with interrelated cues. *Umea Psychological Reports*. 96, 10.
- Ashbaugh, D. (1999). Quantitative-qualitative friction ridge analysis: an introduction to basic and advanced ridgeology, Boca Raton, FL, CRC Press.
- Association of Chief Police Officers (ACPO) (2012). Harnessing Science and Innovation for Forensic Investigation in Policing: Live-time forensics.
- Baber, C. & Butler, M. (2012). Expertise in crime scene examination: comparing search strategies of expert and novice crime scene examiners in simulated crime scenes. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 54 (3), 413–424.
- Banks, W. P. (1970) Signal detection and human memory. *Psychological Bulletin*. 74, 81 – 86.
- Baranski, J. & Petrusic, W. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics*. 55 (4), 412-428.
- Batha, K. & Carroll, M. (2007). Metacognitive training aids decision making. *Australian Journal of Psychology*. 59 (2), 64–69.
- Bektas, S., Bahadir, B., Kandemir, N. O., Barut, F., Gul, A. E. & Ozdamur, S. O. (2009). Intraobserver and interobserver variability in Fuhrman grading systems for conventional renal cell carcinoma. *Koahsiung Journal of Medical Science*. 25, 596-600.



- Biedermann, A., Bozza, S. & Taroni, F. (2008). Decision theoretic properties of forensic identification: Underlying logic and argumentative implications. *Forensic Science International*. 177 (2–3), 120-132.
- Biedermann, A., Garbolino, P. & Taroni, F. (2013). The subjectivist interpretation of probability and the problem of individualisation in forensic science. *Science & Justice*. 53 (2), 192-200.
- Beier, M. E. & Oswald, F. L. (2007). Is cognitive ability a liability? A critique and future research agenda on skilled performance. *Journal of Experimental Psychology: Applied*. 18 (4), 331-345.
- Beilock, S. L. & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 33, 983-998.
- Berbaum, K. S., Dorfman, D. D. & Franken, E. A. (1989). Measuring observer performance by ROC analysis: indications and complications. *Investigative Radiology*. 24, 228 – 233.
- Bleay, S. (2014). Informal interview, carried out by the author on behalf of the Forensic Science Special Interest Group, Home Office CAST, Sandridge, Hertfordshire.
- Bless, H. & Schwarz, N. (2010). Chapter Six: Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model 1st ed., Elsevier Inc. Available at: [http://dx.doi.org/10.1016/S0065-2601\(10\)42006-7](http://dx.doi.org/10.1016/S0065-2601(10)42006-7)
- Bowman, V. (Ed). (1998 (revised 2002, 2004, 2009, 2014)) Manual of fingerprint development techniques, 2<sup>nd</sup> edition. ISBN 1 85893 9720. London: Home Office.
- Brunswik, E. (1952). The conceptual framework of psychology. *International Encyclopaedia of Unified Science*, 1, (10). Chicago: The University of Chicago Press
- Buckleton, J. S., Evett, I. & Weir, B. (2014). Letter to editor in response to editorial by Risinger et al. *Science and Justice*. 54 (6), 510.
- Budowle, B., Buscaglia, J. & Schwartz Perlman, R. (2006). Review of the scientific basis for friction ridge comparisons as a means of identification: committee findings and recommendations. *Forensic Science Communications*. 8 (1).
- Busemeyer, J.R. & Townsend, J.T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*. 100 (3), 432–59.
- Busey, T. & Loftus, G. (2007). Cognitive science and the law. *Trends in Cognitive Sciences*. 11 (3), 111-117.
- Butler, M. (2014) Using eye-tracking, head-mounted camera technology and verbal protocol analysis as a methodology to better understand Volume Crime Scene Investigator practice. Unpublished PhD Thesis. Teesside University.
- Bjorkman, M., Juslin, M. & Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception and Psychophysics*. 54 (1), 75-81.

- Champod, C. (1995). Edmund Locard – numerical standards and ‘probable’ identifications. *Journal of forensic identification*. 45 (2), 136 – 163.
- Champod, C., Lennard, C., Margot, P. & Stoilovic, M. (2004). Fingerprints and other ridge skin impressions. Boca Raton: CRC Press.
- Champod, C. (2008). Fingerprint examination: towards more transparency. *Law, Probability and Risk*. 7, 111-118.
- Champod, C. (2014). Research focused mainly on bias will paralyse forensic science. *Science & Justice*. 54 (2), 107-109.
- Charlton, D. (2013). Standards to avoid bias in fingerprint examination? Are such standards doomed to be based on fiscal expediency? *Journal of Applied Research in Memory and Cognition*. 2, 71-72.
- Charlton, D. (2006). Forensic expertise and self-trust, forensic update (division of forensic psychology). *The British Psychological Society*. 87, 14-19.
- Charlton, D., Fraser-Mackenzie, P. & Dror, I. E. (2010). Emotional experiences and motivating factors associated with fingerprint analysis. *Journal of Forensic Sciences*. 55 (2), 385-393.
- Charlton, D. (2014). Informal interview, carried out by the author on behalf of the Forensic Science Special Interest Group, Surrey and Sussex Police Fingerprint Bureau, Brighton.
- Chesters, M.S. (1992). Human visual perception and ROC methodology in medical imaging. *Physics in medicine and biology*. 37, 1433–1476.
- Chi, M. T. H. (2006). Two approaches to the study of experts’ characteristics. In K. A. Ericsson et al (Ed.), *The Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press. 21-30.
- Chisum, W. J. & Turvey, B. (2000). Evidence dynamics: Locard’s exchange principle and crime reconstruction. *Journal of Behavioural Profiling*. 1 (1), 1-15.
- Christensen, A. M., Crowder, C. M., Ousley, S. D. & Houck, M. M. (2014). Error and its meaning in forensic science. *Journal of forensic sciences*. 59 (1), 123–6.
- Cole, S. A. (2004). Grandfathering evidence: fingerprint admissibility rulings from Jennings to Llera Plaza and back again. *American Criminal Law Review*. 41, 1189-1276.
- Cole, S. A. (2005). More than zero: accounting for error in latent fingerprint identification. *The Journal of Criminal Law and Criminology*. 95 (3), 985-1078.
- Cole, S. A. (2006). The prevalence and potential causes of wrongful conviction by fingerprint evidence. *Golden Gate University Law Review*. 37, 39-105.
- Cole, S. A. (2014). Individualization is dead, long live individualization! Reforms of reporting practices for fingerprint analysis in the United States. *Law, Probability, and Risk*. 13 (2), 117-150.
- Cooke, M. W. & Jinks, S. (1999). Does the Manchester triage system detect the critically ill? *Journal of accident & emergency medicine*. 16 (3), 179–81.

- Corbin Dwyer, S. & Buckle, J. L. (2009). The space between: on being an insider-outsider in qualitative research. *International Journal of Qualitative Methods*. 8 (1), 54-63.
- Craig, A. (1987). Signal detection theory and probability matching apply to vigilance. *Human Factors*. 29, 645 – 652.
- Crispino, F., Ribaux, O., Houck, M. & Margot, P. (2011). Forensic Science – a true science? *Australian Journal of Forensic Science*. 43 (2-3), 157-176.
- Dale, G. & Arnell, K. M. (2014). Lost in the forest, stuck in the trees: Dispositional global/local bias is resistant to exposure to high and low spatial frequencies. *PLoS ONE*. 9 (7). e98625. doi:10.1371/journal.pone.0098625
- Dane, E., Rockmann, K. W. & Pratt, M. G. (2012). When should I trust my gut? Linking domain expertise to intuitive decision-making effectiveness. *Organizational Behavior and Human Decision Processes*. 119 (2), 187–194.
- Dawes, R. M. (1980). Confidence in intellectual vs. confidence in perceptual judgements. In Lantermann, E. D. & Feger, H. (Eds.), *Similarity and choice: Papers in honour of Clyde Coombs* (pp 327 – 345). Hans Huber, Bern.
- Daubert v. Merrell Dow Pharmaceuticals, Inc. 509, US, 579 (1993).
- Dror, I. E., Peron, A. E., Hind, S. L. & Charlton, D. (2005). When emotions get the better of us: the effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology*. 19 (6), 799-809.
- Dror, I. E., Charlton, D. & Peron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*. 156, 74-78.
- Dror, I. E. & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science & Justice*. 51 (4), 204-208.
- Dror, I. E., Champod, C., Langenburg, G., Charlton, D., Hunt, H. & Rosenthal, R. (2011). Cognitive issues in fingerprint analysis: inter and intra expert consistency and the effect of a ‘target’ comparison. *Forensic Science International*. 208, 10-17.
- Dror, I. E. (2009). How can Francis Bacon help forensic science? The four idols of human biases. *Jurimetrics: The Journal of Law, Science, and Technology*. 50 (1), 93-110.
- Dror, I. E. (2012). Expectations, contextual information, and other cognitive influences in forensic laboratories. *Science and Justice*. 52, 132.
- Dror, I. E. (2013). What is (or will be) happening to the cognitive abilities of forensic experts in the new technological age. *Journal of Forensic Sciences*. 58 (2), 563.
- Dror, I.E. (2016). A Hierarchy of Expert Performance (HEP). *Journal of Applied Research in Memory and Cognition*. 5 (2), 121-127.

Earwaker, H. (2014). Fingerprint evidence processes, stakeholders and influences. Forensic Science Special Interest Group. Available from: <https://connect.innovateuk.org/web/forensics/article-view/-/blogs/uk-fingerprinting-process>

Earwaker, H. C., Morgan, R. M., Harris, J. L. & Hall, L. J. (2015). Fingerprint submission decision-making within a UK fingerprint laboratory: Do experts get the marks that they need? *Science and Justice*. 55 (4), 239 – 247.

Egli, N. M., Champod, C. & Margot, P. (2007). Evidence evaluation in fingerprint comparison and automated fingerprint identification systems – modelling within finger variability. *Forensic Science International*. 167 (2-3), 189-195.

Egli, N. M. & Champod, C. (2014). Evidence evaluation in fingerprint comparison and automated fingerprint identification systems – modelling between finger variability. *Forensic Science International*. 235, 86 – 101.

European Network of Forensic Science Institutes (ENSFI) (2015), Best practice manual for fingerprint examination. ENSFI-BPM-FIN-01.

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

Eubanks, J. L., Killeen, P. R. (1983). An application of signal detection theory to air combat training. *Human Factors*. 25, 449 – 456.

Evens, D. (2012). Risk intelligence: how to live with uncertainty, London: Atlantic Books.

Evetts, I. W. & Williams, R. L. (1996). A review of the sixteen point fingerprint standard in England and Wales. *Journal of Forensic Identification*. 46 (1), 49-73.

Expert Working Group on Human Factors in Latent Print Analysis. (2012). Latent print examination and human factors: improving practice through a systems approach. The report of the Expert Working Group on Human Factors in Latent Prints Analysis.

Fieldhouse, S. J. (2009). Consistency and reproducibility in the deposition and evaluation of latent fingerprints, contributing to an investigation into the effectiveness of a portable cyanoacrylate fuming system and aluminium powder for latent fingerprint development (PhD thesis) Staffordshire University.

Fieldhouse, S. J. & Gwinnett, C. (2016). The design and implementation of a proficiency test for assessors of fingerprint quality, to facilitate collaborative practise in fingerprint research. *Science & Justice*. 56, 231-240.

Fischhoff, B. (1982). Debiasing. In: D. Kahneman, P. Slovic, A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp422-444), Cambridge, England: Cambridge University Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist*. 34, 906–911.

Forensic Science Special Interest Group. (2014). Forensic Science Special Interest Group terms of reference. Available from: <https://connect.innovateuk.org/web/forensics/who-we-are> Accessed: 15 October 2014.

Forensic Science Regulator (2014). Terms of reference of the Fingerprint Quality Standards Specialist Working Group. Available from: <https://www.gov.uk/government/organisations/forensic-science-regulator/about/our-governance#fingerprint-quality-specialist-group> Accessed: 15 October 2014.

Forensic Science Regulator (2016). Codes of practice and conduct, Issue 3. The Stationary Office, London.

Forensic Science Regulator (2016a). Draft fingermark development and image capture codes of conduct. Consultation document. Available from: <https://www.gov.uk/government/consultations/fingermark-development-and-image-capture-codes-of-practice>. Accessed: 23 September 2016.

Förster, J., Liberman, N. & Kuschel, S. (2008). The effect of global versus local processing styles on assimilation versus contrast in social judgment. *Journal of personality and social psychology*. 94 (4), 579–599.

Found, B. (2014). Deciphering the human condition: the rise of cognitive forensics. *Australian Journal of Forensic Sciences*. 47 (4), 1-16.

Found, B. & Ganas, J. (2013). The management of domain irrelevant context information in forensic handwriting examination casework. *Science & Justice*. 53 (2), 154-158.

Fraser-Mackenzie, P. A. F., Dror, I. E. & Wertheim, K. (2013). Cognitive and contextual influences in determination of latent fingerprint suitability for identification judgments. *Science and Justice*. 53, 144-153.

Frese, M. & Keith, N. (2015). Action Errors, Error Management, and Learning in Organizations. *Annual Review of Psychology*. 66 (1), 661–687.

Frese, M., Brodbeck, F., Heinbokel, T., Mooser, C., Schleiffenbaum, E. & Thiemann, P. (1991). Errors in training computer skills: on the positive function of errors. *Human computer interaction*. 6, 77-93.

Garb, H. N. (1986). The appropriateness of confidence ratings in clinical judgement. *Journal of Clinical Psychology*. 42, 190-197.

Gao, Z., Flevaris, A. V., Robertson, L. C. & Bentin, S. (2011). Priming global and local processing of composite faces: revisiting the processing-bias effect on face perception. *Attention, perception & psychophysics*. 73 (5), 1477–1486.

Garbolino, P. & Taroni, F. (2002). Evaluation of scientific evidence using Bayesian networks. *Forensic Science International*. 125 (2–3), 149-155.

Garcia-Retamero, R. & Dhimi, M.K. (2009). Take-the-best in expert – novice decision strategies for residential burglary. *Psychonomic Bulletin & Review*. 16 (1), 163–169.

- Garrett, B. L. (2013). Blinded criminal justice. *Journal of Applied Research in Memory and Cognition*. 2 (1), 73-75.
- Geiselman, R. E., Haight, N. & Kimata, L. G., (1984). Context effects on the perceived physical attractiveness of faces. *Journal of Experimental Social Psychology*. 20 (5), 409–424.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic decision making. *The Annual Review of Psychology*. 62, 451-482.
- Gittelsohn, S., Bozza, S., Biedermann A. & Taroni, F. (2013). Decision-theoretic reflections on processing a fingerprint. *Forensic Science International*. 226 (1–3) e42-e47.
- Gimming, D., Huguet, P., Caverni, J. P. & Cury, F. (2006). Choking under pressure and working memory capacity: When performance pressure reduced fluid intelligence. *Psychometric Bulletin & Review*. 13, 1005-1010.
- Goodacre, S. W., Gilet, M., Harris, R. D. & Houlihan, K. P. G. (1999). Consistency of retrospective triage decisions as a standardised instrument for audit. *Journal of accident & emergency medicine*. 16 (5), 322–4.
- Gonzalez-Vallejo, C. (2002). Making Trade-Offs: A Probabilistic and Context-Sensitive Model of Choice Behavior. *Psychological Review*, 109 (1), 137–155.
- Gregory, R. L. (1970) *The intelligent eye*. London: Weidenfeld & Nicolson
- Guenther, C. L. & Alicke, M. D. (2008). Self enhancement and belief perseverance. *Journal of experimental psychology*. 44 (3), 706 – 712.
- Haber, L. & Haber, R. N. (2008). Experimental or scientific expertise. *Law, Probability and Risk*. 7, 143-150.
- Hall, L. J. & Player, H. (2008). Will the introduction of an emotional context affect fingerprint analysis and decision-making? *Forensic Science International*. 181, 36-39.
- Hall, L. (2014). Informal interview, carried out by the author on behalf of the Forensic Science Special Interest Group. New Scotland Yard, London.
- Hammond, K. R., Hirsch, C. J. & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review*. 71 (6), 438–456.
- Hammond, K. R. & Stewart, T. R. eds. (2001) *The essential Brunswik: Beginnings, explications, applications*. Oxford University Press.
- Hardman, D. (2009). *Judgement and Decision Making: Psychological Perspectives*. Chichester: BPS Blackwell.
- Harvey, N. (1997). Confidence in judgement. *Trends in Cognitive Psychology*. 1 (2), 78-82.

- Heath, C. & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*. 4, 5-28.
- Herr, P. M., Sherman, S. J. & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology*. 19 (4), 323–340.
- Hochberg, L. (2014). Metacognition and Decision-Making Style in Clinical Narratives. PhD Thesis. Rochester Institute of Technology.
- Houck, M. M. (2013). Intellectual infrastructure: a modest critique of forensic science. *Science and Justice*. 53, 1.
- House of Commons Science and Technology Committee (2013). Forensic Science, Volume 1, London: The Stationery Office Limited.
- Humphreys, J. D., Porter, G. & Bell, M. (2008). The quantification of fingerprint quality using a relative contrast index, *Forensic Science International*. 178, 46–53.
- Jayaprakash, P. T. (2013). Practical relevance of pattern uniqueness in forensic science. *Forensic Science International*. 231 (1-3), 403 – 416.
- Joseph, G. M. & Patel, V. L. (1990). Domain knowledge and hypothesis generation in diagnostic reasoning. *Medical Decision Making*. 10, 31-46.
- Kahneman, D. & Tversky, A. (1982). Variants of uncertainty. In: Kahneman, D., Slovic, P., & Tversky, A. (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp509-520), Cambridge, England: Cambridge University Press.
- Kahneman, D., Slovic, P. & Tversky, A. (Eds). (1982). *Judgement under uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Kahneman, D. & Klein, G. (2009) Conditions for intuitive expertise: A failure to disagree. *American Psychologist*. 64, 515–26.
- Kahneman, D. & Miller, D.T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*. 93 (2), 136–153.
- Kanchan, T. & Chattopadhyay, S. (2006). Distribution of fingerprint patterns among medical students. *Academy of Forensic Medicine*. 28 (2), 971–973.
- Karelaia, N. & Hogarth, R. M. (2008) Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*. 134 (3), 404–26.
- Kassin, S. M., Dror, I. E. & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*. 2 (1), 42-52.
- Kaye, D. H. (2003). Questioning a courtroom proof of the uniqueness of fingerprints. *International Statistical Review*. 71 (3), 521-533.

- Keren, G. (1988). On the ability of monitoring non-veridical perceptions and uncertain knowledge: Some calibration. *Acta Psychologica*. 67, 95-119.
- Kerstholt, J., Eikelbooma, A., Dijkmanb, T., Stoel, R., Hermsen, R. & Van Leuven, B. (2010). Does suggestive information cause a confirmation bias in bullet comparisons? *Forensic Science International*. 198 (1-3), 138–142.
- Kerstholt, J.H., Paashuis, R. & Sjerps, M. (2007). Shoe print examinations: Effects of expectation, complexity and experience. *Forensic Science International*. 165(1), 30–34.
- Keuth, H. (2005). *The philosophy of Karl Popper*, Cambridge: Cambridge University Press.
- Koehler, J. J. (2008). A welcome exchange on the scientific status of fingerprinting. *Law, Probability and Risk*. 7, 85-86.
- Kruger, A. N. & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*. 199 (2), 254 – 284.
- Kruglanski, A. W. & Webster, D. M. (1996). Motivated closing of the mind: seizing and freezing. *Physiological Reviews*. 103 (2), 263-283.
- Krupinski, E.A. (2011). The role of perception in imaging: Past and future. *Seminars in Nuclear Medicine*. 41 (6), 392–400.
- Krupinski, E. A., Tillack, A. A., Richter, L., Henderson, J. t., Bhattacharyya, A. K., Scott, K. M., Graham, A. R., Descour, M. R., Davis, J. R. & Weinstein, R. S. (2006). Eye movement study and human performance using telepathology virtual slides: Implications for medical education and differences with experience. *Human Pathology*. 37, 1543-1556.
- Kundel, H. L., Nodine, C. F. & Carmody, D. (1978). Visual scanning pattern recognition and decision making in pulmonary nodule detection. *Invest. Radiol*. 13, 175-181.
- Kundel H. L., Nodine C. F., & Krupinski E. A. (1989). Searching for lung nodules. Visual dwell indicates locations of false-positive and false-negative decisions. *Invest. Radiol*. 24, 472-478.
- Laber, T., Kish, P., Taylor, M., Owens, G., Osborne, N. & Curren, J. (2014). Reliability Assessment of Current Methods in Bloodstain Pattern Analysis, NIJ 2010-DN-BX-K213. Available from: <https://www.ncjrs.gov/pdffiles1/nij/grants/247180.pdf>. Accessed: 5 March 2015.
- Lagden, A. (2014). Informal interview, carried out by the author on behalf of the Forensic Science Special Interest Group at Harperly Hall, Durham.
- The Law Commission (2011). *Expert evidence in criminal proceedings in England and Wales*, London: The Stationery Office.
- Langenburg, G. M. (2004). Pilot study: a statistical analysis of the ACE-V methodology—analysis stage. *Journal of Forensic Identification*. 54, 64–79.



Langenburg, G. (2009). A performance study of the ACE-V process: a pilot study to measure the accuracy, precision, reproducibility, repeatability, and biasability of conclusions resulting from the ACE-V process. *Journal of Forensic Identification*. 59 (2), 219-257.

Langenburg, G. (2012). A critical analysis and study of the ACE-V process, (unpublished doctoral dissertation), Universite´ de Lausanne, Lausanne. Available from: [http://www.unil.ch/files/live//sites/esc/files/shared/Langenburg\\_Thesis\\_Critical\\_Analysis\\_of\\_ACE-V\\_2012.pdf](http://www.unil.ch/files/live//sites/esc/files/shared/Langenburg_Thesis_Critical_Analysis_of_ACE-V_2012.pdf).

Langenburg, G, Champod, C. & Genessay, T. (2012). Informing the judgments of fingerprint analysts using quality metric and statistical assessment tools. *Forensic Science International*. 219 (1-3), 183-198.

Leadbetter, M. (2007). Letter to the editor. *Fingerprint Whorld*. 129 (33), 231.

Liberman V. & Tversky, A. (1993) On the evaluation of probability judgments: Calibration, resolution, and monotonicity. *Psychological Bulletin*. 114, 162-173.

Mackenzie, R. (2011). Witness statement of Robert Mackenzie. The Fingerprint Inquiry Scotland. 17-18. Available from: [http://www.thefingerprintinquiryscotland.org.uk/inquiry/files/FI\\_0046.pdf](http://www.thefingerprintinquiryscotland.org.uk/inquiry/files/FI_0046.pdf) Accessed: 18 April 2013.

Mangels, J, A., Butterfield, B., Lamb, L., Good, C. & Dweck, C. S. (2006). Why do beliefs about intelligence influence learning success? A social cognitive model. *Social, Cognitive and Affective Neuroscience*. 1, 75-86.

Marsh, N. (2015). Verbal advice in relation to image resizing in Adobe Photoshop, Metropolitan Police Service, Lambeth, London.

Mello-Thoms, C., Mello, C. A. & Medvedeva, O. (2011). Characterizing virtual slide exploration through the use of "search maps." *Proceeds of SPIE Med Imaging*, 7966:79660Q.

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*. 18 (4), 361 - 376.

Mikaelyan, A. & Bigun, J. (2012). Ground truth and evaluation for latent fingerprint matching. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE. 83-88. Available from: <http://ieeexplore.ieee.org/document/6239220/>. Accessed: 13 August 2016.

Mnookin, J. L. (2008). The validity of latent fingerprint identification: confessions of a fingerprint moderate. *Law, Probability and Risk*. 7, 127-142.

Morgan, R. M., Allen, E., King, T. & Bull, P. A. (2014). The spatial and temporal distribution of pollen in a room: Forensic implications. *Science and Justice*. 54 (1), 49-56.

Murch, R., Abbott, A., Fox, E., Hsiao, M. & Budowle, B. (2012). Establishing the quantitative basis for sufficiency thresholds and metrics for friction ridge pattern detail and the foundation for a standard. Final Technical Report, Distributed by US Department of Justice.

Murphy, A. H. & Winkler, R. L. (1977). Can weather forecasters formulate reliable probability forecasts of precipitation and temperature? *National Weather Digest*. 2, 2-9.

Nakhaeizadeh, S., Dror, I. E. & Morgan, R. M. (2014). Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. *Science & Justice*. 54 (3), 208 - 214.

Nanakorn, S., Kutanan, W. & Chusilp, K. (2013). An exploration of fingerprint patterns and their concordance among Thai adolescents. *Chiang Mai Journal of Science*. 40 (3), 332–343.

National Research Council Committee on Identifying the Needs of the Forensic Science Community (2009). Strengthening Forensic Science in the United States: A Path Forward.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*. 9 (3), 353–383.

Lewis, M. B., Mills, C., Hills, P. J. & Weston, N. (2009). Navon letters affect face learning and face retrieval. *Experimental Psychology*. 56(4), 258–264.

Neumann, C., Mateos Garcia, I., Langenburg, G., & Kostroski, J. (2011). Operational benefits and challenges of the use of fingerprint statistical models: a field study. *Forensic Science International*. 212, 32-46.

Neumann, C., Evett, I. W., Skerrett, J. E., Mateos Garcia, I. (2012). Quantitative assessment of evidential weight for a fingerprint comparison. Part II: A generalisation to take account of general pattern. *Forensic Science International*. 214 (1-3), 195-199.

Neumann, C., Champod, C., Yoo, M., Genessay, T. & Langenburg, G. (2015). Quantifying the weight of fingerprint evidence through the spatial relationship, directions and types of minutiae observed on fingermarks. *Forensic science international*. 248, 154–71

Neumann, C., Armstrong, D.E., & Wu, T. (2016). Determination of AFIS “Sufficiency” in Friction Ridge Examination. *Forensic Science International*. 263, 114–125.

Newell, B.R. & Shanks, D.R. (2014). Unconscious influences on decision making: a critical review. *The Behavioral and brain sciences*. 37 (1), 1–19.

Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review*. 84 (3), 231-259.

Osborne, N. K. P., Woods, S., Kieser, J. & Zajac, R. (2014). Does contextual information bias bitemark comparisons? *Science & Justice*. 54 (4), 267-273.

Paese, P. W. & Sniezek, J. A. (1991). Influences on the appropriateness of confidence in judgement: Practice, effort, information, and decision-making. *Organizational behavior and human decision processes*. 48 (1). 100-130.

- Page, M., Taylor, J. & Blenkin, M. (2012). Context Effects and Observer Bias—Implications for Forensic Odontology. *Journal of Forensic Sciences*. 57 (1), 108-112.
- Pallant, J. (2007). *SPSS Survival Manual; a step by step guide to data analysis using SPSS for Windows*. 3<sup>rd</sup> ed. Open University Press, Maidenhead.
- Parker, A. M. & Fischhoff, B. (2005). Decision making competence: External validation through an individual-differences approach. *Journal of Behavioural Decision Making*. 18, 1-27.
- Phillips, V. L., Saks, M. J. & Peterson, J. L. (2001). The application of signal detection theory to decision-making in forensic science. *Journal of Forensic Sciences*. 46 (2), 294 - 308.
- Plous, S. (1993). *The Psychology of Judgement and Decision Making*, McGraw-Hill, New York.
- Potchen, J. E., Cooper, T. G., Sierra, A. E., Aben, G. R., Potchen, M. J., Potter, M. G. & Siebert, J. E. (2000). Measuring performance in chest radiography. *Radiology*. 217, 456-459.
- Pulsifer, D. P. (2013). An objective fingerprint quality- grading system. *Forensic Science International*. 231 (1-3), 204 – 207.
- Reason, J. (1990). *Human Error*. Cambridge University Press, New York.
- Risinger, D. M., Thompson, W. C., Jamieson, A., Koppl, R., Kornfield, I., Krane, D., Mnookin, L. J., Rosenthal, R., Saks, M. J. & Zabell, S. L. (2014) Regarding Champod, editorial: Research focused mainly on bias will paralyse forensic science. *Science & Justice*. <http://dx.doi.org/10.1016/j.scijus.2014.06.002>
- Robin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*. 83 (5), 1281-1302.
- Saks, M. J. (2010). Forensic Identification: from a faith based ‘science’ to a scientific science. *Forensic Science International*. 201, 14-17.
- Saks, M. J. & Koehler, J. J. (2008). The individualization fallacy in forensic science evidence. *Vanderbilt Law Review*. 61 (1), 199-219.
- Sauer, J. D., Brewer, N. & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*. 137, 528–547.
- Sauer, J. D., Brewer, N., & Weber, N. (2012). Using confidence ratings to identify a target among foils. *Journal of Applied Research in Memory and Cognition*. 1, 80–88.
- Schiffer, B., & Champod, C. (2007). The potential (negative) influence of observational biases at the analysis stage of fingerprint individualisation. *Forensic Science International*. 167, 116-120.
- Schwarz, N., Strack, F., Mai, H. P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*. 55 (1), 2–23.

- Scott, K.R., Morgan, R. M., Jones, V. J. & Cameron, N. G. (2014). The transferability of diatoms to clothing and the methods appropriate for their collection and analysis in forensic geoscience. *Forensic Science International*. 421, 127-137.
- Sears, V. G., Bleay, S. M., Bandey, H. L. & Bowman, V. J. (2012). A methodology for fingerprint research. *Science and Justice*. 52 (2012), 145–160.
- Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST). (2011). Standards for examining friction ridge impressions and resulting conclusions (latent/tenprint), version 1.0.
- Shepherd, D. A. & Zacharakis, A. (2002). Venture capitalists' expertise: A call for research into decision aids and cognitive feedback. *Journal of Business Venturing*. 17 (1), 1-20.
- Silverman, B. (2011). Research and Development in Forensic Science: a review. Home Office Science. London.
- Silverman, D. (2011). Interpreting Qualitative data 4<sup>th</sup> edition, Chapter 3.3 Grounded Theory, SAGE, London
- Skala, D. (2008). Overconfidence in psychology and finance – an interdisciplinary literature review. *Bank i Kredyt*. 4, 33-50.
- Somoza, E. & Massman, D. (1991). 'Biological markers' and psychiatric diagnosis: risk benefit balancing using ROC analysis. *Biological Psychiatry*. (1991) 29, 811 – 826.
- Stambouli, H., El Bouri, A. & Tijana, N. (2015). Occurrence of fingerprint patterns in the Moroccan population. *Canadian Society of Forensic Science Journal*. 48 (4), 160–166.
- Stewart, N., Chater, N. & Brown, G. D. A., (2006). Decision by sampling. *Cognitive Psychology*. 53(1), 1–26.
- Stewart, N., (2009). Decision by sampling: the role of the decision environment in risky choice. *Quarterly journal of experimental psychology*. 62 (6), 1041–1062.
- Stoel, R. D., Dror, I. E. & Miller, L. S. (2014). Bias among forensic document examiners: Still a need for procedural changes. *Australian Journal of Forensic Sciences*. 46, 91–97.
- Stow, K. (2014). Informal interview, carried out by the author on behalf of the Forensic Science Special Interest Group, Derbyshire.
- Swets, J. A. & Birdsall, T. G. (1967). Deferred decision in human signal detection: a preliminary experiment. *Perception Psychophysics*. 2, 15 – 28.
- Swets, J. A. (1992). The science of choosing the right decision threshold in high stakes diagnostics. *The American Psychologist*. 47, 522 – 532.

- Swofford, H., Steffan, S., Warner, G., Bridge, C. & Salyards, J. (2013). Inter and intra- examiner variation in the detection of friction ridge skin minutiae. *Journal of Forensic Identification*. 63 (5), 553-571.
- Tetlock, P. E. & Kim, J. I. (1987). Accountability and judgement processes in a personality prediction tasks. *Journal of Personality and Social Psychology*. 52, 700-709.
- Thompson, W. C. (2010). Beyond bad apples: analyzing the role of forensic science in wrongful convictions. *Southwestern University Law Review*. 1, 971–994.
- Thompson, M. B., Tangen, J. M. & McCarthy, D. J. (2013). Expertise in fingerprint identification. *Journal of Forensic Sciences*. 58 (6), 1519-1530.
- Tobe, S., Dennany, L. & Vennemann, M. (2015). An assessment of the objectivity of sperm scoring. *Forensic Science International*. 251, 83-86.
- Triplett, M. (2013). Errors in forensics: Cause(s) and solutions. *Journal of Applied Research in Memory and Cognition*. 2(1), 63-64.
- Tully, G. (2015). Forensic Science Regulator Annual Report 2014 – 2015. United Kingdom Home Office.
- Tversky, A., & Koehler, D. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*. 101 (4), 547–567.
- Ulery, B. T., Hicklin, R. A., Kiebusinski, G. I., Roberts, M. A., & Buscaglia, J. (2013). Understanding the sufficiency of information for latent fingerprint value determinations. *Forensic Science International*. 230, 99-106.
- Ulery, B. T., Hicklin, R. A., Roberts, M. A. & Buscaglia, J. (2014). Measuring what latent print examiners consider sufficient information for individualisation determinations. *PLoS ONE*. 9 (11), e110179.
- Ulery, B. T., Hicklin, R. A., Roberts, M. A. & Buscaglia, J. (2014a). Changes in latent fingerprint examiners' markup between Analysis and Comparison. *Forensic Science International*. 247, 54–61.
- Ulery, B. T., Hicklin, R. A., Roberts, M. A. & Buscaglia, J. (2016). Inter-examiner variation of minutiae mark-up on latent fingerprints. *Forensic Science International*. 264, 89-99.
- Versteeg, C. H., Sanderink, G. C. H., Lobach, S. R. & Van der Stelt, P. F. (1998). Reduction in size of digital images: does it lead to less detectability or loss of diagnostic information? *Dentomaxillofacial Radiology*. 27, 93 – 96.
- Vokey, J. R., Tangen, J. M. & Cole, S. A. (2009). On the preliminary psychophysics of fingerprint identification. *The Quarterly Journal of Experimental Psychology*. 62 (5), 1023-1040.

Wedell, D. H., Parducci, A. & Geiselman, R. E. (1987). A formal analysis of ratings of physical attractiveness: Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology*. 23 (3), 230–249.

Wells, G. L., Olson, E. A. & Charman, S. D. (2002). The confidence of eyewitnesses in their identifications from lineups. *Current Directions in Psychological Science*. 11, 151–154.

Wells, G. L., Wilford, M. M. & Smalarz, L. (2013). Forensic science testing: the forensic filler-control method for controlling contextual bias, estimating error rates, and calibrating analysts' reports. *Journal of Applied Research in Memory and Cognition*. 2, 53-55.

Winman, A. & Juslin, P. (1993). Calibration of sensory and cognitive judgements: Two different accounts. *Scandinavian Journal of Psychology*. 34, 135-148.

Wood, A. M., Brown, G. D. A. & Maltby, J. (2012). Social norm influences on evaluations of the risks associated with alcohol consumption: applying the rank-based decision by sampling model to health judgments. *Alcohol and alcoholism*. 47 (1), 57–62.

Zakay, D. Ellis, S. & Shevsky, M. (2004). Outcome value and early warning indicators as determinations of willingness to learn from experience. *Experimental Psychology*. 51, 150-157.

## List of Appendices

### Appendix A – Chapter 2

Appendix A.1. Roles within fingermark enhancement

Appendix A.2. Roles within fingermark recovery

Appendix A.3. Roles within fingerprint comparison

### Appendix B – Chapter 3

Appendix B.1. Earwaker, H., Morgan, R. M., Harris, A. J. & Hall, L. J. (2015). Fingermark submission decision-making within a UK fingerprint laboratory: Do experts get the marks that they need? *Science and Justice*, 55 (4), 239-247.

Appendix B.2. Fingerprint examiner fingermark assessment sheets

Appendix B.3. Profile of experimental fingermarks

Appendix B.4. Decoy mark details

Appendix B.5. Serious crime instruction sheet

Appendix B.6. Serious crime results sheet

Appendix B.7. Volume crime instruction sheet

Appendix B.8. Volume crime results sheet

Appendix B.9. Participant information sheet

Appendix B.10. Participant consent form

Appendix B.11. Laboratory manager interview schedule

Appendix B.12. Data files relating to Objectives 3.1- 3.4

### Appendix C – Chapter 4

Appendix C.1. Decision rationale data

### Appendix D – Chapter 5

Appendix D.1. General instruction sheet

Appendix D.2. Experimental pack distribution instructions

Appendix D.3. Timescale record sheet

Appendix D.4.	Participant instruction sheet
Appendix D.5.	Participant information sheet
Appendix D.6.	Participant consent form
Appendix D.7.	Demographic information sheet
Appendix D.8.	Response sheet
Appendix D.9.	Minutiae count data
Appendix E – Chapter 6	
Appendix E.1.	Images of mock exhibits detailing the origin and location of each mark
Appendix E.2.	Final experimental exhibit images
Appendix E.3.	Written instructions for the distribution of experimental packs
Appendix E.4.	Record of pack distribution
Appendix E.5.	Participant instruction sheet
Appendix E.6.	Participant information sheet
Appendix E.7.	Informed consent form
Appendix E.8.	Demographic information form
Appendix E.9.	Participant instructions
Appendix E.10.	Verbal instructions to be provided to participants
Appendix E.11.	Minutiae context data