ERRARE HUMANUM EST? DECISION-MAKING PROCESSES, COGNITIVE BIAS, AND MOTIVATION OF FINGERPRINT EXAMINERS

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by
Francisco Valente Gonçalves
Department of Criminology
University of Leicester

2020

ERRARE HUMANUM EST?
DECISION-MAKING PROCESSES, COGNITIVE BIAS, AND MOTIVATION OF FINGERPRINT EXAMINERS

Francisco Valente Gonçalves

The attention that academia and forensic agencies have been giving to the decision-making process of fingerprint examiners has increased in recent years. This is largely due to the increasing awareness that experts' decisions can suffer from cognitive bias, promoting erroneous decisions. Even though fingerprint examiners are expected to not commit errors deliberately, it is of interest to understand in detail the influences (i.e. impact on performance) that external factors such as contextual information have on their decision-making process.

This thesis begins by investigating how a group of experts (n = 41) and another of laypeople (n = 57) differ in their accuracy and response time regarding the influences of different contextual information. Although experts showed higher levels of accuracy, findings suggest that both laypeople and experts have a tendency to suffer from similar types of cognitive bias associated with the same types of contextual information promoting lower accuracy and higher response times. It seems that different types of contextual information have different types of influence in experts' performance. Hence, this thesis analysed in a second study the accuracy and response times of 67 fingerprint examiners, from 15 forensic bureaus, based in 9 countries, during trials that simulated the Verification phase of the ACE-V process. Results showed significant differences between specific types of contextual information when compared with control trials, challenging guidelines that suggest fingerprint examiners should work within a full blind setting. Due to the diverse sample in this study, it was possible to explore differences regarding the different levels of experience, methodological approaches currently in use (i.e. numerical approach versus holistic approach) and the accreditation standards that forensic bureaus had. Findings provided insight that can be used in future developments for methodologies, accreditation guidelines, and training for fingerprint examiners.

Finally, a qualitative study was conducted in which 42 fingerprint examiners were interviewed regarding their opinions about contextual information influences and the methodologies within the ACE-V process. Results retrieved from interviews shed light on aspects of experts' work, allowing a better understanding of the motivation and the level of cognitive enjoyment (assessed by the level of Need for Cognition) as well as the specific types of contextual information which may not influence experts' performance negatively, but rather be a motivational factor for their work.

Acknowledgements

All of one's work is surely due to effort, patience, resilience, and creativity to solve challenges. However, one needs not to embrace all the challenges alone. Indeed, the journey of completing a PhD is much of sharing and embracing support. And for that, one should be thankful.

Thank you to Professor Lisa Smith and Dr Doug Barrett. For being always supportive. For accepting my ideas and challenging them in a way I could see gaps and make them better day after day. For treating me as a colleague rather than a student in a lower rank. For making me feel listened and encouraged to embrace innovative ways to seek solutions for my questions. And to support me during personal issues I had along this journey. For making me feel I could count on them. Always. Thank you.

I will never forget the way I was received in all fingerprint bureaus I visited. I cannot forget any of the fingerprint examiners that I spoke with, shared ideas, and challenged to conduct my experimental activities. All of them were extremely humble and collaborative. All of them were always easy to access and talk to and to receive feedback regarding new ways of work. I cannot express my sincere thanks to all fingerprint examiners that collaborated with me.

Thank you to all the people I met within the academic community, namely from the Marie Curie Alumni Association, EuroScience and the Portuguese Association of Researcher and Students in the UK. It was a pleasure to be able to meet people that are brilliant minds, and who I exchanged ideas and insights related to all sorts of topics, from mental health to astrophysics!

My consideration also needs to address the efforts that the European Commission has been doing to keep Research in Europe to be carried out. These are times of change, but I rather expect the future to be defined more by the people, and less by politics.

To all my friends that are always there. In good moments, but especially in the sad too. Friendship is notably one of the blessings that mankind has. Sharing it with the group of people I am surrounded is something one needs not an explanation. But instead to be grateful. Thank you.

To my family. My pillars in life. Those who taught me to be humble enough to say with a smile “I don’t know, please explain to me”. That taught me to accept ideas and ask questions when needed. Life is not happy all the time, but you have made me skilled enough to overcome obstacles, even when they seem impossible to do so.

To Leicester as a city. I never thought I could like so much the UK apart from London. Thank you to open my doors to the real UK, the real ale, the proper pie, and the rainy days. I cannot say I loved it, but being apart from it now, I can say that sometimes, I miss it. A lot.

A final note to also thank my new team at the Forensic and Integrity Services at EY. This new journey has been a great way to put into action all of the knowledge I gathered throughout my PhD. Thank you for giving me the time I needed to complete this thesis and to support me always. It has been great to work in such a place.

“Intelligence is the ability to adapt to change”

Stephen Hawking

CONTENTS

LIST OF TABLES AND FIGURES

Tables

Figures

RECOGNITIONS, PUBLICATIONS & PRESENTATIONS RESULTING FROM THIS THESIS

Recognitions

EUROSCIENCE & ELSEVIER (€500)
Top 3 2020 European Young Researcher Award (EYRA)

MARIE CURIE ALUMNI ASSOCIATION (€1500)
Social Impact Award

MARIE SKŁODOWSKA-CURIE ACTIONS | EUROPEAN COMMISSION
Bridging Career Paths Award

MARIE CURIE ALUMNI ASSOCIATION (€400)
Micro Grant for research

POSTGRAD RESEARCH FESTIVAL UNIVERSITY OF LEICESTER (£350)
Best poster overall and public engagement prize

POSTGRAD RESEARCH FESTIVAL UNIVERSITY OF LEICESTER (£250)
Best poster of the social sciences college

POSTGRAD SCHOOL UNIVERSITY OF LEICESTER (£367)
Micro Research Grant Award

MARIE CURIE ALUMNI ASSOCIATION (€400)
Micro Grant for research

MARIE CURIE ALUMNI ASSOCIATION UK CHAPTER
Outstanding Research Engagement Pitch, 3rd place

PORTUGUESE INSTITUTE OF LEGAL MEDICINE AND FORENSIC SCIENCES
Best Oral Presentation, 1st place

ELTU // UNIVERSITY OF LEICESTER (£100)
Outstanding Research Presentation, 1st place

Publications and presentations

Journal articles

Gonçalves, F. V. (2017). Differences between fingerprint agencies across numerical and holistic approaches. *Fingerprint Whorld, Chartered Society of Forensic Sciences.* 2, 18-21

Liu, S., Zhongliang, Mi, Z., Gonçalves, F. V., Wu, J. & Ayers, K. (2017). Development of Quality Control Systems for Fingerprint Comparison Processes. Journal of Forensic Science and Medicine. 3(2), 49-54

Posters

Gonçalves, F. V., Smyth, A., Smith, L. & Barrett, D. (2017). Decision Making in Forensic Science. Postgrad Festival at University of Leicester. Leicester, UK

Gonçalves, F. V., Smith, L. & Barrett, D. (2017). Beyond The Context: Decision Making in Forensic Science. Postgrad Festival at University of Leicester. Leicester, UK

Gonçalves, F. V., Smith, L. & Barrett, D. (2017). Decision-Making in Fingerprint Comparisons. Marie Curie Annual Conference. Salamanca, Spain

Gonçalves, F. V., Smith, L. & Barrett, D. (2016). Decision Making in Fingerprint Expertise: Contextual Information and Motivation. VIII Simpósio Forense IPEBJ. Ribeirão Preto, Brazil

Gonçalves, F. V., Smith, L. & Barrett, D. (2016). Think Outside the Box: External Factors on the ACE-V Methodology. 68th American Academy of Forensic Sciences Metting. Las Vegas, U.S.

Gonçalves, F. V., Smith, L. & Barrett, D. (2015). Let's Fill the Gaps! Getting fingerprint examiners into the scene. CPD day of Surrey/Sussex Police. Brighton

Gonçalves, F. V., Smith, L. & Barrett, D. (2015). Errare Humanum Est? Cognitive contaminations in fingerprint examiners' performance: Attitudes and improvements. ELTU Research Festival held at the University of Leicester. Leicester, UK

Gonçalves, F. V., Smith, L. & Barrett, D. (2015). How does fingerprint analysis work? Open day with the theme of Richard III discovery at the University of Leicester. Leicester, UK

Gonçalves, F. V., Smith, L. & Barrett, D. (2015). Decision-Making in fingerprint analysis. Open day with the theme of DNA Fingerprinting discovery by Sir Alec Jeffreys at the University of Leicester. Leicester, UK

Gonçalves, F. V., Smith, L. & Barrett, D. (2015). It is more than just errors: A quantitative and qualitative approach to cognitive contaminations within the ACE-V methodology. NIST Symposium on Standards and Errors Measurement. Washington DC, U.S.

Gonçalves, F. V., Smith, L. & Barrett, D. (2015). A psychological perspective within forensic pattern recognition disciplines. European Association of Psychology and Law Conference. Nuremberg, Germany

Gonçalves, F. V., Smith, L. & Barrett, D. (2015). A criminological perspective concerning misguided decision in forensic sciences. European Society of Criminology. Porto, Portugal

Oral communications

Gonçalves, F. V., Smith, L. & Barrett, D. (2018). Errare Humanum Est: A Journey through Forensic Sciences. INTREPID Forensics Conference. University of Leicester & Chartered Society of Forensic Sciences

Gonçalves, F. V. (2017). Forensic Sciences and Artificial Intelligence: Reflections for the future. codeX Research Group. Stanford University, U.S.

Smyth, A., Gonçalves, F. V. & Jensen, S. (2017). New Forensic Technology: A Validation Study. IAI EU Conference, Amsterdam, The Netherlands

Smyth, A., Gonçalves, F. V. & Jensen, S. (2017). New Forensic Technology: A Validation Study. 21st Triennial Meeting of the International Association of Forensic Sciences. Toronto, Canada

Gonçalves, F. V., Smith, L. & Barrett, D. (2017). To know or not to know. What should be the context. Breaking New Grounds in Psychology and Law. European Association of Psychology and Law. Mechelen, Belgium

Gonçalves, F. V., Smith, L. & Barrett, D. (2017). Differences between Holistic and Numerical Approach on the ACE-V Process. The Chartered Society of Forensic Sciences, Fingerprint Division. Manchester, UK

Gonçalves, F. V., Smith, L. & Barrett, D. (2016). The Latent Imbroglio: Pattern Recognition with Fingerprint Examiners, Contextual Information and Motivation. 2nd Conference of the European Division of the International Association for Identification. INTERPOL Headquarters, Lyon, France

Gonçalves, F. V., Smith, L. & Barrett, D. (2016). Fingerprint Experts' Motivation When Carrying Out the ACE-V Process. Australian and New Zealand Society of Forensic Sciences. Auckland, New Zealand

Gonçalves, F. V. (2016). Uma luz ao fundo do túnel? Estandardização no direito e nas ciências forenses. III Conferência Instituto Nacional de Medicina Legal e Ciências Forenses. Coimbra

Gonçalves, F. V., Smith, L. & Barrett, D. (2015). Do we really need to know it all? What fingerprint experts need to know during the ACE-V process. The Chartered Society of Forensic Sciences postgraduate symposium. Manchester, UK

Gonçalves, F. V., Smith, L. & Barrett, D. (2015). Internal and external variables within the ACE-V process. 1st International Association of Identification, European division. Leicester, UK

Gonçalves, F. V., Smith, L. & Barrett, D. (2014). Decision-Making within fingerprint analysis. University of Lausanne Forensic Sciences Summer School. Lausanne, Switzerland

1. LITERATURE REVIEW

1.1. INTRODUCTION

On an everyday basis, people are constantly making decisions. Just as an example, in the first hour of the morning one can observe several actions that derive from decisions made by people such as when to get out of bed, what is the perfect temperature for a shower, what clothes to wear, what to eat, and how to travel to work. Although the number of decisions that one makes in the first hours of a day may vary from person to person, most of those decisions may be taken unconsciously, i.e., automatically. Some other actions ask for one's reflection and consciousness within the decision-making process. In each day, an average adult makes approximately 35,000 decisions in contrast to a child that only makes 3000 conscious decisions (Sahakian & Labuzetta, 2013). The set of actions that are considered automatic processes depend on a certain level of experience or repetition, whereas more considered actions have a degree of novelty, importance or complexity. It is also important to point out that automatic decisions are possible to make within a short – almost null – time-frame, whereas complex decisions may demand time to think, analyse and make the decision.

Automatic decisions may lack consciousness of potential consequences. Complex decisions, on the other hand, require further reflection. It is also important to recognise that the decision-making process is influenced by a wide set of variables. Humans are influenced by external variables that may lead them to make decisions with less accuracy. Some of these external influences act within people's internal cognition and are so-called cognitive biases (Tversky & Kahnemann, 1974). One of the challenges in the field of decision-making processes is the need to make people aware of those inaccurate decisions that usually are conducted unconsciously and automatically. Successfully alerting individuals to the importance of their decisions will depend on the impact that decisions have on people's lives. For instance, choosing a bad movie to watch has very little impact on one's life. However, selecting a candidate to be the primary suspect of a serious crime is going to impact not only the individual's life but also the criminal justice system as a whole.

The focus of this thesis is the decision-making processes within a field where consequences have a significant impact on people's lives: The Criminal Justice System and the forensic sciences. The main objective of this thesis is to address some of the gaps within the decision-making literature that exists in a specific field of forensic sciences, namely fingerprint identifications, and to discuss how procedures can be improved and mitigate potential errors.

Throughout this thesis, a literature review is presented in the first chapter to review the existing knowledge on decision-making processes from a general perspective and then focused on the decision-making processes in forensic sciences, specifically in fingerprint examinations. Within the main topic, there are sections covering topics associated with errors such as cognitive biases and experts' motivation. The literature review chapter is then followed by the methodology and empirical chapters where the experimental work is presented using a mixed-method methodology.

The originality of this thesis is highlighted in the three empirical chapters (Chapters 3, 4 and 5) which identified aspects related to examiners' performance of high relevance for better understanding the decision-making processes, the influences of cognitive bias, and the motivation that fingerprint examiners have throughout their work. Chapter 3 describes a comparison between laypeople and fingerprint examiners regarding the influences of different types of contextual information, challenging the notion that contextual information influences only in a negative way the performance of experts. Chapter 4 includes further analysis of the influences that different contextual information has within experts' performance. The inferences that were possible from the results in this chapter are of interest due to the fact that it has been claimed previously that contextual information is a source of cognitive bias, however, results in this chapter showed that not all contextual information promoted erroneous decisions. The sample of participants in this study was also of interest since fingerprint examiners that participated in this study worked in 15 fingerprint bureaus (based in nine countries) where different methodological approaches (numerical approach versus holistic approach) were in place as well as different quality standards. A novel insight that this study provided was the application of two psychometric surveys, one that assessed the Need for Cognition and another that assessed the Extrinsic and Intrinsic Motivation of

fingerprint examiners, allowing a reflection on these two concepts and the relationship between performance and the effects of contextual information. Finally, in Chapter 5, contextual information being a potential source of cognitive bias was also addressed, however, using a qualitative methodology. In this study, fingerprint experts were interviewed, and their responses were thematically analysed, and insights related to their views and attitudes were described. This study is an important contribution to the literature as there is a limited amount of research that encouraged fingerprint examiners to provide their opinions on important topics such as the relationship between contextual information and performance as well as other topics (e.g. blind setting, motivation, limitations of the job). The overall discussion presented in Chapter 6 presents an interesting and novel way to observe results. Results from Chapters 3, 4 and 5 were combined in order to have a holistic view of fingerprint examiners' procedures and suggest recommendations that forensic bureaus can implement as well as exploring some explanations for the effects of cognitive bias within the practice.

1.2. DECISION-MAKING PROCESSES

In 1981, Daniel Kahnemann and Amos Tversky published one of the papers which would revolutionise the field of studying decision-making processes. They claimed decision making to be a set of dynamic actions in which there are at least two phases: Editing and Evaluating (Tversky & Kahnemann, 1981). Referring to the editing phase, the authors explained it as the initial step where people internally create a representation of the problem, putting together, with the problem itself, the features which interact with it such as actions, outcomes, and contingencies. On the other hand, evaluating is composed of the weight [positive or negative] that decision-makers assign to a given decision. The authors conducted a study (Tversky & Kahnemann, 1981) where they presented participants with a problem to solve in a country [in their study Sweden] that needed to fight a dangerous disease [AIDS]. To observe how people perceived the problem, two scenarios were presented. In both scenarios, one of two actions needed to be chosen. In the first scenario, Program A suggested that 400 people would die whereas in Program B there was 1/3 chance that nobody would die and 2/3 probability that 600 people would die. Scenario 2 presented, mathematically, the same outcome but reworded. The choice of Program A would suggest that 200 people would die and Program B suggested

the possibility of 1/3 chance that nobody dies and 2/3 chance that nobody survives. Even though the outcomes in the programs that were presented were the same, the results indicated that participants changed their decision based on the framing they were presented within each scenario. The authors described scenario 1 as having a *negative framing* whereas scenario 2 was considered a *positive framing* scenario.

People begin by framing the scenario where the decision will impact. This framing moment is constrained by two conditions: (1) the characteristics and guidelines or habits decision-makers follow and (2) the formulation of the problem itself, i.e. its evaluation. Evaluation is composed of two variables: (a) value and (b) probability. Kahnemann and Tversky (1981) suggested that the process of evaluation becomes more complex as the perception of losses and gains is greater. Thus, the possibilities that decision-makers have at the beginning of the decision-making process are transformed from possible alternatives to actual gains and losses, affecting the estimated value that people attribute to those possible alternatives.

Kerstholt and Raaijmakers (1997) reflected on the variety of variables one can experience during decision making, such as time pressure, the complexity of the task itself or even the environment in which the decision is made. They invested their efforts in describing an approach to observe the process of decision making that is very relevant to this thesis, the decision-making process of a dynamic task, which is relevant for fingerprint examiners' work due to the fact that these examiners work within a dynamic context where variables can change and where the process that is followed (known as ACE-V) could be considered as a dynamic process, i.e. it has some differences on its application for instance due to the type of laboratory where it is being applied, the type of internal motivation the examiner has to carry their work, access to technology, amongst other variables.

Dynamic tasks are described as a set of possible choices that are conducted within environments that can exogenously change or that are made as a result of previous decisions, and where those decisions are sequentially linked to each other through their effects. Due to the dynamism of those decisions, actions at specific times will directly or indirectly influence future actions (Gonzalez, Fakhari & Busemeyer, 2017). Kerstholt and Raaijmakers (1997) stated the need to use different strategies to make decisions due to the change and transformation of some variables such as

the context. As illustrations of their rationale, the authors described the type of decision a firefighter makes when controlling fire, or perhaps the diagnosis made by a physician. In these examples, variables such as fire's direction and intensity can change due to weather conditions, and patients' diagnoses can also change due to their responses, the possibility to provide accurate information to physicians or even the detail of the clinical record physicians can gather.

Kerstholt and Raaijmakers (1997) suggested four characteristics of dynamic tasks that can influence decision makers' responses. However, when aware of those, decisions can be as accurate as in perfect conditions. The first one was related to the change over time that a decision is confronted with. For instance, in the example of a physician's diagnosis, the time passing might promote deterioration of the patient's condition, and due to that, physicians need to change their course of action.

The second factor presented was the availability of feedback. Most important, feedback can be used to test accuracy of actions and to adjust those when needed (Kleinmuntz, 1985). Feedback is used to advance the accuracy of the decision-making system, even incrementally, by making small changes that adjust performance and lead to greater transformations. In this respect, there are two types of changes that can be made due to feedback (Kleinmuntz, 1985). Individuals that apply action-oriented strategies will apply actions only, observe the effects of the outcomes within the system under control and then apply changes. On the other hand, decision-makers that use judgment-oriented strategies will first try to reduce the uncertainty that is grasping the root of potential flaws by gathering more information and then make the decisions that are needed.

Thirdly, dynamic tasks are carried out within several interdependent decisions. This means that these types of tasks require multiple decisions that affect the system itself. Decision-makers need to ensure that they have an accurate mental representation with enough quality (i.e. enough information) related to the relations within the system as well as all parameters that affect the system. If they do not possess an accurate mental representation of the system, decision-makers may fall into a position where they cannot anticipate future outcomes. A fourth and final characteristic of dynamic tasks is the possibility to create experimental tasks based on dynamic tasks. By doing that, decision-makers will be able to learn from the

different responses that are possible to give at a certain moment due to a change that has happened. Once decision-makers learn from these experimental tasks, they will be able to apply better strategies to solve problems that are presented in real-life decisions.

Decision-makers select the strategy that works best most of the time for them, and people utilise short-cuts when needed (Payne, Bettman and Johnson, 1993). Hence, variables such as time might affect individuals' performance since people adopt adaptive processes to achieve at least "*two goals: attaining an accurate performance and investing minimal effort*" (Kerstholt and Raaijmakers, 1997, p.214). However, Maule and Edland (1997) emphasize a different perspective. Even though they followed the suggestion that decision-makers can pursue shortcuts due to time constraints, they suggested that performance can be better within time pressure as individuals will assess the same amount of information in shorter time periods, making them rely on short-cut strategies. Aronson (1999) suggested an interesting phenomenon linked with decision-making processes that make use of shortcuts. This is the concept of *cognitive misers* that the author claimed to be a way to conserve cognitive energy. Aiming to overcome complex problems, individuals simplify those into simple tasks. However, Aronson warned, people tend to solve complex problems by ignoring some information to reduce their cognitive load. He also stated how this kind of process can be the root of possible errors and biases, once people start bypassing steps within the tasks. By doing that, decision-makers may fall into selecting the wrong information to consider and process, leading them to possible flawed decisions.

Variables within the context of a decision-making system can constrain accuracy and performance (Kahneman, 2011b; Kerstholt and Raaijmakers, 1997; Kahneman, Slovic & Tversky, 1982). However, instead of expecting to find only poorer outcomes due to the variability of some of the factors within the decision-making system, some research presented another perspective besides limited outcomes. Research also suggested that time pressure, for instance, could be a variable that promotes conservative decisions, i.e. decisions that are made without too much confidence, and due to the lack of confidence, an alternative and more cautious option is taken (Ben-Zur & Breznitz, 1981). These kinds of decisions entirely depend on the decision-makers' perception of risk and the weight they put on the

decision itself, which according to Kaplan, Wanshula, and Zanna (1993) is affected by the sources of information one has access to. Huber and Kunz (2007) observed their participants making risky decisions under time pressure. The study was carried out with 40 participants who were asked to make decisions within three different scenarios. The first scenario illustrated an environmental problem regarding a turtle species, the second scenario was related to a virus infection epidemic situation. A third scenario, adapted from a previous study, was used as a warm-up task prior to the experimental conditions. Scenarios given to participants presented two possible solutions, a risky solution, and a non-risky solution. Participants could ask for more information during the task if wanted, however, time constraints were imposed as well. Results in both experimental scenarios showed a large effect size of time pressure. Under time pressure, the search for more information was reduced in general and participants made more conservative decisions ($p<.001$). Regarding time pressures, there were significant differences in the type of information that participants requested. Authors found that under time pressure, 73% of participants asked more questions about negative consequences, whereas when the time was not strongly pressuring respondents, the search for positive consequences was greater by 27%. Within time pressure, respondents became more conservative by seeking for negative consequences of their decisions. Although accuracy could not be measured as a variable in this experiment, the authors suggested that with time pressure participants tried to gather more information regarding negative consequences, which can be translated into an attempt to challenge the option itself, and therefore making respondents more conservative when they did not have enough confidence in their decision.

1.3. EXPERTS' DECISION-MAKING PROCESSES

Society tends to believe that experts are, in general, more accurate than novices (Wynne, 1996). Laypeople are usually more inclined to rely on what is known as experts' advice or opinions (Torngren & Montgomery, 2010). That tends to happen due to experts focus on variables that laypeople tend to give less importance or trust (Siegrist, Keller, Kastenholz, Frey & Wiek, 2007). Although experts are considered to have more knowledge in their fields than laypeople, they need to have what Shanteau and Gaeth (1983) called the *expert image* to be accepted by lay-people. However, this *expert image* may sometimes be misinterpreted with some

characteristics or attitudes such as arrogance, feeling of self-importance or overconfidence being observed in fields like medicine (Shanteau & Gaeth, 1983; Golde, 1970).

The definition of expertise can be associated with the early 80's, where Gruber (1982) made a note on the "*interest of human beings at their best*" (p. 248). At this time, experts' skills assessments were still at an embryonic stage. However, the assessment of expertise started earlier from research on education (Sosniak, 2006; Bloom, 1985, 1982). Concepts such as talented, giftedness, genius, prodigious and others were the focus of research that, eventually, led to current research on experts' assessment and expertise development. However, it is not completely clear the right nomenclature one should use due to different views on this concept.

There are at least two well-known perspectives that have been exploring ways to describe and understand expertise: (a) The Expert Performance Approach (EPA) (Ericsson, Charness, Feltovich & Hoffman, 2006) and (b) The Performance-Based Approach (PBA) (Shanteau, 1992b). These perspectives seem to have some differences in their views. Whereas the EPA sees expertise within an *absolute* view (Chi, 2006), i.e. without a continuum of potential evolution for the experts, the PBA defines expertise within a *relativistic* perspective (Weiss & Shanteau, 2014), i.e. within a flexible and continuous approach of how expertise is developed. What is meant by this difference is that EPA suggests the role of the expert and the evolution of its expertise to be somehow established within structured parameters. The PBA claims expertise to be a phenomenon that one achieves within a flexible organisation. Both views articulate the idea that an expert will follow a scalable line to achieve the status of being considered an expert. However, the organisation to achieve that level is rather different between approaches.

There are four major differences between both perspectives: (1) what is the definition of expert, (2) what one needs to be considered an expert, (3) what are the levels/types of expertise and (4) what are the ways to assess expertise. Within EPA, an expert is someone with more knowledge about a certain field than non-experts [lay-people], who achieved his/her level of expertise within a structured way (Ericsson & Smith, 1991). EPA also presumes that domain-general reasoning abilities of experts can be found in non-experts and the logical way to assess the

differences between experts and non-experts is to determine the differences in their knowledge.

2.2.1. PBA approach

Shanteau (1988), advocating for the PBA approach, described ten core characteristics that experts need to have in order to create and maintain what the author suggested is the "*self-presentation of an expert*" to society:

Characteristic / Personality trait	**Explanation**
Perceptual/Attention	Ability to extract information that lay-people overlook or do not even pay attention to.
Relevance assessment	Capability to differentiate relevant from irrelevant information.
Simplify complexity	Turning a complex problem in something simpler and organised.
Good communication	Capability of communicating effectively their expertise. Capability to persuade/convince others.
Handling adversities	Ability to face adversities and keep making effective decisions. (e.g. working under stress).
Outward confidence	Capacity to believe themselves and their decisions.
Content knowledge	Ability to be up-to-date with the latest developments in a specific field.
Creativity	Capacity to develop creative strategies to solve difficult problems
Verbal inarticulate/automaticity	Inability to express themselves in clear ways for laypeople to understand, using vague sentences such as "that's just the way it is".

Table 1 – Core characteristics of an expert (Shanteau, 1988, pp. 209-211)

Within a relativistic view, PBA suggests that judgement is the core of an expert's responsibility (Weiss & Shanteau, 2014). Academics that recommend this approach for the study of expertise see individuals and their expertise within a continuum of capabilities instead of "*distinctions as cut-points*" (Weiss & Shanteau, 2014, p. 1) as seen within an absolute approach like EPA. To assess expertise the PBA makes use of a metaphor suggested by Weiss and Shanteau (2003), *the expert judge*. The authors claimed intra-reliability and inter-reliability to be concepts that may promote confusion when assessing expertise due to some aspects such as linguistics on a consensus of terminology:

> "*The confusion has arisen because consensus is the basis for terminology. Constructs, such as the defining characteristics of a disease, must be shared by the linguistic community that employs*

> *them. Doctors need to agree on what is meant by a term such as myocardial infarction. (…) The judgment depends on more than merely knowing what the diagnostic category entails. Perhaps a crucial symptom is hard to detect, so that only someone with superior vision or sense of smell notices it. Whether the judgment is correct cannot be determined by agreement among judges.*" (Weiss & Shanteau, 2003, p. 4)

Due to this potential confusion in terminology, instead of aiming for consensus and assessing consensus between experts, the PBA suggestion is to understand expertise as a ratio between *discrimination* and *inconsistency* to assess intra and inter reliability of experts. The model succeeded Cochran's work (Cochran, 1943) and his suggestion that a ratio used as an assessment of quality should allow perceived differences consistently between assessments of similar stimuli. This way, Weiss and Shanteau suggested then the Cochran-Weiss-Shanteau index as:

CWS = Discrimination / Inconsistency

It is possible to denote the emphasis that Weiss and Shanteau (2003) wanted to transmit on the evaluation of expertise. Both perspectives (EPA and PBA) observe the types of expertise and their assessment in their way. For the work conducted in this thesis, the PBA approach, which places judgement at the core of the expertise will be adopted as the selected model to assess expertise. Within the model, *discrimination* is related to the different assessments that different experts make within certain stimuli set. *Consistency* is associated with the moment when the *expert-judges* assess the same stimuli, in different events, within similar ratings.

To outline this model, PBA advocates re-analysed a study by Skånér, Strender and Bring (1998). From a sample of 27 physicians, Weiss and Shanteau (2003) picked four physicians and analysed their rates for discrimination and consistency during diagnoses in heart failure situations. The four physicians that were selected assessed five patients each. Physicians were asked to diagnose each patient two times, although in the second assessment they did not know that cases were being repeated. Physicians were asked to do their assessments within a Likert scale between *totally unlikely* to observe heart failure to *certain* to observe heart failure. Across the four

physicians that were re-assessed, figure 1 shows that only one physician (doctor #18) was significantly more discriminating and consistent. Doctor #8 was rated with a good score only in the discriminating competence. Doctor #16 was consistent but had similar assessments for all patients making this participant's scoring in discrimination low, and finally, the last individual (doctor #23) was neither discriminating or consistent.

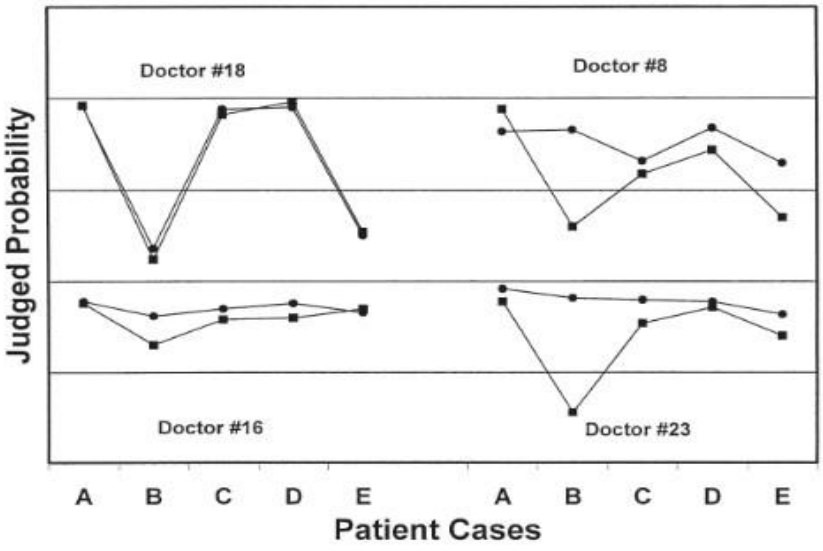

Figure 1. Physicians' judgment (Weiss & Shanteau, 2003)

Within the rationale of the CWS index, only doctor #18 showed high expertise levels due to the demonstrated high discrimination and low inconsistency. Doctor #8 presented moderate discrimination and high inconsistency. The third physician (doctor #16) that was assessed presented low discrimination and low inconsistency and finally doctor #23 had low discrimination and high inconsistency. From the CWS viewpoint, the authors suggest that an individual that is insensitive to the stimuli, i.e., responds randomly to similar stimuli will have a typical CWS value = 1. Decision-makers within a field of expertise that show high competences in their field (as doctor #18 showed) will present *effective discrimination* and a low inconsistency. Effective discrimination means that individuals will evaluate stimulus according to the changes that exist within the set of stimuli. Low inconsistency implies that decision-makers will assess similar stimuli within the same level of rating. Weiss and Shanteau (2003) suggested the use of mean squares as potential dispersion variances to assess CWS values for the

model proposed, which this thesis will also use in some of the statistical analysis throughout the empirical work.

This work conducted by Weiss and Shanteau (2003) suggested that some professionals that can be considered experts may, at some point, show differences in their outcomes for the same assessment. The rationale questions expertise to be considered as something which suggests that only talented people will perform well. The PBA and the use of an index such as the CWS index suggest that due to potential variations within the performance, either experts or laypeople may suffer from flaws and misguided decision-making, namely due to biases. This can also be observed in Kahnemann's words, *"(...) Furthermore, there is much evidence that experts are not immune to the cognitive illusions that affect other people (...)"* (Kahneman, 1991, p. 144).

Although the characteristics of experts may sound possible to find in non-experts, other variables make expert judgements essential in specific circumstances. As an example, Burgman and colleagues (2011) mentioned that expert judgements are "*attractive when time and resources are stretched, and essential where data are inadequate, circumstances are unique, or extrapolations are required for novel, future and uncertain situations*" (p. 1). Hence, it could be expected that experts who had similar levels of expertise, would provide similar outcomes to the same questions/challenges, i.e., that they are reliable between them. However, this was not true in all the cases as observed in previous research where experts showed inter-reliability (Stewart, Roebber & Bosart, 1997; Einhorn, 1974).

Shanteau (1988) suggested three types of expertise based on individuals' skills. Each type of expertise presents two types of experts accordingly. The first type of expertise is the *analysis* type which includes within it the *cognitive experts* and the *perceptual experts*. Cognitive experts use their sensory skills to perceive differences not relevant to non-experts, for instance, livestock judges. Perceptual experts possess unique problem-solving skills used to see relations that usually are not perceived by non-experts (e.g. auditors). The second type of expertise the author mentioned in his work was the type of expertise based on *knowledge*, which encompasses the *knowledge experts*, who base their answers in large sets of information (e.g. academics) and the *diagnostic experts* (e.g. physicians) that use a

limited amount of information. Lastly, the third type of expertise focuses on the level of involvement that experts show, which can be *advice* (e.g. consultants) or *action* itself (e.g. business managers).

Weiss and Shanteau (2003) updated the previous organisation for a structure that presents four types of expertise based on the skills experts have and in the type of evaluation they make throughout their decisions. They considered that experts may have a type of expertise (1) based in their quantitative and qualitative evaluation (expert judgement), (2) based in a projection (expert prediction), (3) based in the communication the expert has and shows (expert instruction) or (4) based in the execution (expert performance). All four types of expertise suggested by PBA's advocates kept the main rationale of this approach, that expertise is observed in a relativistic way, meaning that expertise is not a rigid concept, but rather a set of skills that are under development as opposed to the EPA approach.

2.2.2. EPA approach

EPA's absolute approach suggested expertise as a concept that can be studied when the aim is to evaluate the individuals who are exceptional (Chi, 2006). Being a less flexible view of expertise when compared with the relative approach, EPA splits the types of expertise by knowledge domain instead of skills that are used/needed and the way individuals perform tasks when they are asked to demonstrate their abilities in a specific field. Related to this rationale, Hoffman (1998, p. 83) presented seven types of expertise, based on the level of expertise that one is attributed to.

Within the EPA approach, there are seven levels of expertise that were described (Hoffman, 1998), as table 2 presents:

Level of expertise	**Description**
naïve	An individual considered ignorant about a defined subject or domain.
Novices	Individuals that are just starting and did have limited exposure to the domain of expertise
Initiate	Individuals who have been through an initiation procedure and have started already their introductory instruction

Apprentice	Credited to people who are learning about a certain domain after the introductory phase
Journeyman	Individuals who can carry out some of the tasks within a certain domain have when these tasks are unsupervised
Expert	A distinguished or brilliant journeyman, who is highly regarded by peers, whose judgements and performance are often reliable
Master	A master should be considered as an extension of a great expert who is also qualified to teach those at lower levels. Traditionally, a master level of expertise is going to be a level where one finds individuals who set standards and regulations as well as ideals for a specific domain

Table 2 – Levels of expertise within EPA approach (Hoffman, 1998)

Regarding the organisation proposed by the EPA approach and its suggestion of assessing individuals with proficiency tasks, it is important to discuss the differences in how both approaches observe expertise. On one hand, it seems important to identify different types of expertise and have these organised by categories. However, it may seem somehow limited to have these categories static. The reason for this can be found in research where inter-reliability within expertise seems to increase with experience. Ettenson, Shanteau, and Krogstad (1987) observed responses from 32 auditing cases carried out by 11 professional auditors and 11 accounting students, which had the same number of significant factors to support participants' decisions. Professional auditors retrieved only a limited amount of information to achieve the same decision as students (i.e. relevant information retrieving), suggesting that experience made them faster (mentioned as *relevance assessment* previously (Shanteau, 1988) as well as having significant inter-reliance according to their experience.

Lundberg, Forsell, Johansson, and Josefsson (2014) emphasised how experience is important and how it can facilitate the accomplishment of tasks that individuals carry out. The authors used an eye tracker to observe how the information cued by flight controllers was processed, finding that senior controllers needed significantly less time than trainees to picture the whole air traffic flow during what is commonly referred as conflict moments (i.e. moments when flights can crash), however, novices achieved the same outcome as senior controllers. Moertl, Bonaceto, Estes, and Burns (2008) observed the way air traffic controllers assess relevant information and make their decisions when there is the possibility of having changes in their procedures such as conflicts between aeroplanes when landing. In

both groups (experts and novices), one of the most significant factors that the authors claimed these professionals needed to have, as a skill, was anticipated separation, which is the way controllers manage the traffic flow. With this capacity, controllers predicted what they think was going to happen based on judgement, which is grounded in their knowledge and previous experience.

Aiming to better understand experts' performance, Chi (2006) suggested the need to acknowledge not only situations in which experts are extremely good at carrying out their tasks, but also the importance of observing situations and conditions in which experts may demonstrate lower performance than what is expected. With this objective in mind, Chi (2006) presented 13 types of skills that characterise performance. Seven of them were associated with excelling performance, whereas six were related to performance that falls short. Regarding the skills that excel performance, the author mentioned that experts can (1) generate the best solutions faster and more accurately than non-experts, (2) detect and recognise features within problems that novices are not able to do, (3) analyse a problem from a qualitative viewpoint better than non-experts as well as developing its representation by adding domain-specific knowledge, (4) self-monitor their skills, being capable to detect errors and the status of their knowledge, (5) wisely choose better appropriate strategies when compared with novices, (6) use resources in a more opportunistic way than novices to solve a problem, and finally experts (7) control more their cognitive load and retrieve more relevant knowledge on domain-specific fields than novices. On the other hand, experts can fall-short in their expertise due to (1) domain-limitations, i.e. when they need to recall knowledge from other fields they do not have enough knowledge to be considered experts, (2) being overly confident, and achieving accuracy rates similar to novices, (3) glossing over circumstantial details as novices do, (4) dependency on contextual cues, (5) in need to re-adapt during changes because their established knowledge, (6) biases that affect their decision-making processes.

Training is for both perspectives one of the most important phases that individuals complete in order to achieve expert status within a certain field of work. Historically, fields such as medicine and law were domains where professionalization was permitted only for specific and certified professionals (Ericsson, 2006). Laboratory experiments have led the way to offer information

about how individuals acquire their skills by allowing researchers to control the training conditions and observing the effects of independent variables which can be isolated (Proctor & Vu, 2006). This type of laboratory setting allowed research to suggest a view where there is a transference of knowledge within complex tasks (Willingham, Wells, Farrel, Stemwedel, 2000; Willingham, 1999). Healy, Kole, and Bourne (2014) mentioned three forms of task engagement for successful training when individuals conduct complex tasks: Acquisition, retention, and transfer. The authors suggested that for a positive rate of acquisition, there should be feedback scheduled in slots of time. This increases the training speed, which facilitates the identification and correction of errors as well as the promotion of motivation for trainees. Regarding the difficulty of the inputs that the trainer aims to implement, there is an optimal *zone of learnability* that needs to be balanced between what the subject already knows, and he/she is asked to perform more rapidly, with higher accuracy or with new types of information that are beyond the current knowledge of the trainee.

To prevent the trainees from disengaging in the task, and to mitigate potential errors in prolonged tasks, such as being in front of a computer routinely performing the same set of actions, some strategies can be applied, such as splitting the training into small parts or including a cognitive task in the workflow to mitigate against the declining of accuracy (e.g. typing a set of digits and ending the list of numbers alternating between the + or the – key). Secondly, there is the retention of the training which is linked with the durability of the new skills. Healy, Kole, and Bourne (2014) claimed that there are two types of tasks that influence retention. One that affects the power of efficiency and another that does not affect efficiency. Extra tasks that are associated with the main skill that is being learned promote retention, however, if the extra task is not relevant for the learning of the main skill, retention will be negatively affected as well as performance.

Based on the two different approaches (PBA and EPA) that were explored, it is possible to observe some issues that still need some development and uniformization. Firstly, and probably most important, conceptualisation. Although expertise and associated research on this concept started many years ago, there are still some gaps between schools of thought on how to define expertise and organise a methodology to assess it across individuals as well as to categorise the different

types of expertise. Secondly, it is important to understand that in different fields, the way expertise is observed differs due to the context. And finally, it seems that expertise can be assessed through different traits or characteristics, which all people have. Nevertheless, excellent performance is more expected to be observed in individuals that developed their training throughout their experience. However, it will always be possible to observe the event of someone committing an error due to unconscious and automatic decisions. And this phenomenon of committing errors due to cognitive bias occurs not only with people that are considered non-experts but also in experts as will be demonstrated in the following section.

Two approaches to observe and assess expertise were explored. The PBA perspective that suggests a relative and flexible view to assess expertise. For that reason, this approach was chosen to support the understanding of expertise and the decision-making processes that experts make. Nevertheless, the EPA approach also added value to the understanding of expertise as well as the characteristics of an expert. Both perspectives have some points in common, for example, the fact that experts are, in general, more accurate than lay-people. However, the idea that people who are considered experts can have their performance affected similarly to novices or even laypeople in specific moments and due to specific variables, such as bias, may sound threatening to the value that society gives to an expert, specially within fields as the Criminal Justice System. Yet, it is not misleading and erroneous. Indeed, this idea was suggested by some of the most influential authors in the field of decision making (Tversky & Kahneman, 1971; Kahneman, Slovic & Tversky, 1982) as well as disseminated in social media and traditional media (Shariatmadari, 2015) where experts were described as having a level of accuracy that could be negatively affected by some traits that experts tend to have such as overconfidence. Nevertheless, the work that experts do, and the level of experience individuals need to achieve to become an expert, are obviously of greatest importance when compared to lay-people's performance.

It happens that experts, will, in fact, be affected by variables that non-experts are too. One of the potential explanations for that may be the excess of confidence that experts have when performing a task, or even the motivation to carry out a specific task. It is also interesting to note that society seems to give more importance to an expert's error than to an error committed by a lay-person. Even with higher stakes

associated with experts' decisions, it is important to highlight from this section that in terms of accuracy, experts are far more accurate than lay-people in most tasks where their expertise is needed, even though they can be affected by variables such as bias.

1.4. COGNITIVE CONTAMINATIONS, HEURISTICS, AND BIAS

The connotation of the word bias itself relates specifically to activities that may lead to errors, and due to the meaning of this word, this thesis adopts the term *contamination* in favour of bias. To achieve the necessary understanding of this concept within the cognitive processes, this thesis takes into consideration previous research within the field of biased decision-making and related concepts from a general approach, ending with the two types of bias which will be addressed specifically in the section on forensic sciences and that will be explored within the empirical chapters, namely the confirmation bias and the contextual bias.

This chapter started by mentioning that people make a great number of decisions throughout the day. Some of these decisions have consequences that affect one's life in a very small way and with low impact. Nevertheless, other decisions may have a huge impact on individuals' lives. Taking as an example, a document written by a physician regarding a patient's health condition. If this document has any mistake, the possibility of a misguided diagnosis is real. But medicine is not the only field where errors have a great impact on individuals' lives. A product that enters the market with the wrong strategy, a less accurate description of a commercial flight made by an air traffic controller or a forensic report regarding evidence that can link a suspect to a certain crime. These are illustrations of a bigger set of examples that one can think of to illustrate misguided decisions where the stakes are extremely high in terms of consequences. Even though society may claim explanations and scapegoats for terrible actions that led one to commit errors, psychology should not focus its attention on the outcome, but rather on the procedure that led an individual to make a certain decision that progressed into an error. According to Cohen (1993) what psychologists should take into consideration, more than the outcome of a wrong decision, is the process and the associated cognitive activity that were involved in that misguided procedure, such

as, false beliefs, inappropriate priorities, shaky inferences from data or even logical inconsistencies that one is aware of but still ignores.

Kahnemann (1991) described a set of works that were published in the 1950's that were focused on the subjectivity and uncertainty of decision and judgments, characterised by three factors, (i) the critical attitude towards the normative theory of rational belief and choice, (ii) the emphasis on risky choices, and (iii) the preference for analyses that invoke cognitive or psychophysical terms, neglecting some of the emotional and social factors. Observing factors like these, Kahnemann and Tversky (1982) introduced the concept of heuristics. The heuristics and biases approach, according to Kahnemann, had its focus on the previous three factors, and two new ones (p. 142): the first one was related to the "*emphasis on cognitive processes described at an intermediate level (heuristics of judgements, framing option)*", where these were used to predict and explain the diversity of phenomenon in judgment and choice. The second factor was associated with a "*research method that favours simple between-subjects experiments drawing on subjects' life experience*". The term heuristic was used for the first time by George Polya in 1945 when he was trying to demonstrate how mathematicians think and retrieve that kind of reasoning model to teach students (Baron, 2014). Putting the term in a simple way, Baron (2014) explained that a heuristic is a type of rule that has unclear conditions, contrasting for instance with an algorithm that has strict and clear rules to be applied. However, heuristics do not always have negative consequences. Using heuristics may be an intelligence trait, as it is a shortcut that the human brain uses to get a faster answer for a determined problem as demonstrated in Chapter 1 by Aronson (1999), nevertheless decisions based in heuristics are prone to be biased. Tversky and Kahnemann (1974) mentioned that many decisions are based on existent beliefs. The authors illustrated this idea by mentioning that the expected outcomes can be based on the likelihood of uncertain events that did not happen but are very similar to other events such as the election of a specific candidate for the presidency, the market value of a currency in the future or even the guilt of a defendant.

The cognitive contaminations that heuristics lead to can be found in a wide range of events within at least three types of contexts as described by Cohen (1993): (i) assessment of probabilities – where overconfidence in estimating probabilities

either in simple or complex events is required or even when the decision weight is recalled by previous experiences, the data retrieved by valid statistics may seem less important than other anchors that individuals may have as pre-beliefs; (ii) inferences – when people tend to disregard or ignore previous data and assess an event with a single piece of content instead of observing the entire picture, promoting an overestimation either by making random or extreme assessments; (iii) choice – which vary for instance due to the way a given problem is verbally described.

An important paper on heuristics (Tversky & Kahnemann, 1974) presented the three most prominent and important heuristics that led the way for the study of cognitive contaminations such as the contextual and confirmation biases. The first type of heuristic presented was *representativeness*. This heuristic is associated with types of problem-solving that people usually do by using probabilities between a known concept/object versus another not known so well (e.g. the probability of event A being originated by event B or the probability that one specific object A belongs to one specific class B). As mentioned previously, our brain tends to adopt shortcuts and therefore people tend to assume that their *representative idea* of a specific case is more *probable* to happen than using relevant statistical data (Pi, Parisi & Luppi, 2014).

To illustrate intuitive predictions, Kahnemann and Tversky (1973) ran a set of studies noting two classes of a prediction made by individuals based on representativeness: (i) category prediction [given in a nominal form, i.e., predicting the person to be elected, a diagnosis or a future career] and (ii) numerical prediction (given in a specific number/value, for instance, a student's grade or future value of a stock). In one of these studies, the authors conducted a simple, but indeed very important experiment to observe category prediction, where they asked three groups of subjects (Group 1 – base-rate group, N=69; Group 2 – similarity group, N=65; Group 3 – predictive group, N=114) to make estimations regarding the choice of field studies of an individual American first-year graduate student. Groups of participants had been given information on statistics related to first-year graduate choices, however, experimental groups (group 2 and group 3) also received information related to first-year graduate personalities. Group 2 only had a sketch (i.e. limited information) related to the personality of a regular first-year graduate

student. Group 3 in its turn was provided with a full personality assessment provided by a high school psychologist of a regular American first-year graduate. Correlations were significantly different between group 1 and group 2 (F = .97), and non-significant between group 1 and group 3 (F = -.65).

Taking into consideration the significance of the first correlation in favour of the second, the authors claimed that a "direct confirmation of the hypothesis that people predict by representativeness, or similarity" (p. 239) could be assumed, adding as well that this prediction based on representativeness had more weight than for instance the use of statistical data, i.e., subjects in the cited study were not capable of analysing specific and valid data but rather they based their decision almost as a *guess* based on their previous beliefs.

Another study carried out by the same authors (Kahnemann & Tversky, 1973) verified how subjects performed when asked to assign specific values to test representativeness within the numerical prediction. Three groups were asked to predict the grade point average of 10 hypothetical students based on a single percentile score obtained in a specific factor. Each group had different factors that percentiles were associated with (percentile for group 1 was referring to the score on academic achievements; percentile for group 2 referred to the ability to concentrate; group 3 received a percentile that referred to the possibility that students had to appreciate humour).

Group 1 had the information about its percentile associated with an average grade point, however, group 2 and group 3 did not (group 2 – mental concentration and group 3 – sense of humour). There was also additional information provided that referred to correlations between groups 2 and 3 to the average grade point (i.e. students with higher levels of concentration or sense of humour were expected to have higher grades points on average). However, it was mentioned that these correlations were not always valid or even very accurate.

Results of this study showed interesting relationships. Group 1 was expected to reproduce their values, and that was what respondents did. However, group 2 and group 3 predicted grade points average from unrelated variables such as mental concentration and sense of humour. Authors expected group 2 to be less regressive

as the mental concentration test could be assumed to have some reliability to predict grade points average, however, they expected higher variability within group 3 since humour was not assumed to be a valid measure of academic ability. There were no significant differences between the predictions from group 1 and group 2, suggesting that respondents were primed (i.e. biased) when receiving information that was assumed to be less accurate and invalid to predict grade points such as mental concentration.

In the studies described above and others that were summarised in the same paper (Kahnemann & Tversky, 1973) it was assumed that the representativeness hypothesis was supported in the experimental designs, showing that "predictions are no more regressive than evaluations or judgments of similarity" (p. 248). This set of experiments also alludes to the fact that individuals tend to anchor on specific information that is "available prior to the experiment, in the form of stereotypes or expectations" (p. 248), suggesting the potential flaws that will be explored in cognitive contaminations such as the confirmation bias or the contextual bias.

Following the representativeness heuristic, Tversky and Kahnemann (1974) mentioned the *Availability* heuristic, which was described as the phenomenon that individuals experience when they assess how events can happen derived from previous experiences recalled, i.e. the availability that people tend to use their capabilities to dig into their memory and bring previous events to their current reality. Cohen (1993) highlighted that using the availability heuristic could involve the addition of other information within the recalling action, however, individuals tend to take the information they can recall at a first glance to the decision problem they need to answer, instead of creating the space to reflect on the information that is unequivocally needed.

An earlier paper by Tversky and Kahnemann (1973) described this heuristic too. Authors organised the availability heuristic into three types: (i) assessments of availability, (ii) availability for construction and (iii) availability for retrieval. The study asked 152 participants to perform a judgment task where they needed to estimate the frequency of finding within a typical English text, words where the third letter was either K, L, N, R or a V or words that would start with those same consonants. Within the sample, 69% of respondents judged the first position to be

more likely ($p < .001$) for most of the letters in the list even though there was a higher number of words that had the third letter one of the consonants in the list provided. This showed the authors that people tend to judge based on the ease that they will be able to retrieve the information from their mind, rather than on the exactitude of the information.

Finally, *adjustment* and *anchoring* were reported as the last heuristic that leads one to act within a biased way (Tversky & Kahnemann, 1974). The reasoning behind the anchoring process implies that individuals who are making decisions do anchor in a decision first, making then a series of insufficient adjustments before reaching a final answer, which is biased towards the prior anchor (Chapman & Johnson, 2002). Anchoring may also be perceived as the result of overconfidence that an individual has within a certain field (Cohen, 1993) or instead in any given field where knowledge is basic and anchors. This was demonstrated by Tversky and Kahnemann (1974) who carried out a study where they asked participants to estimate the number of African countries that were included in the United Nations. In this study, participants were asked this question after seeing a spinning wheel giving a number between 0 and 100. Individuals were firstly asked whether their estimation was above or below the number in the spinning wheel followed by their specific estimation. Results in this study showed that the number that the spinning wheel gave influenced individuals' estimations, which was demonstrated by significant differences ($p = .001$) between groups that saw the number 10 or the number 65 on the spinning wheel.

The human brain performs an extremely fast recall of an event, and that action will not necessarily be a problem or the start of a potential error. However, there are gaps created by the heuristics processes associated with this type of fast sequencing which may lead to different types of biases such as (a) bias due to the retrievability of instances – observed regarding the size of a class and the ease that the instances of that particular class are retrieved; (b) bias due to the effectiveness of a search set which is observed in the different strategies that the human brain carries in specific searches; (c) biases of imaginability that occur when people need to estimate a certain frequency that is not known, nevertheless can be produced, usually with some biases; (d) illusory correlation which the authors noted in their 1973 paper (Tversky & Kahnemann, 1973) as the frequency of co-occurrence, associated with

the recalling of an item-pair, i.e., the capacity that the brain has to link two events into a single item/fact.

It is clear that heuristics play an important role in the decision-making processes that people do in their everyday lives. As seen above, evidence suggests that the different types of heuristics, such as representativeness, exert a great influence in different fields such as medicine (Gilovich & Savitsky, 2002) regarding the disparity of weights that individuals give to a specific event-based not in science or valid data, but rather on overoptimism (Pi, Parisi & Luppi, 2014). Individuals usually do not recognise the occurrence of an unconscious error, and when they do, often they will try to use alternative reasons to explain their decisions (Tversky & Kahnemann, 1974) as it was demonstrated within the work that judges sometimes do when suffering from this type of bias in their sentencing (Enough & Mussweiller, 2006).

1.4.1. Confirmation Bias

As observed in the previous section, people tend to follow their initial hypotheses, and by doing that, they might experience the phenomenon of anchoring, and potentially fall into what is known as the confirmation bias. The confirmation bias occurs when people search and/or assess information in a way that supports pre-existing beliefs and preconceptions (Nickerson, 1998) to favour hypotheses with unwarranted tenacity and confidence (Klayman, 1995). Oswald and Grosjean (2004) mentioned that the systematic way that information is processed promotes lay people or experts to accept those hypotheses if the final aim is meant to confirm those.

A study carried out by Darley and Gross (1983) showed the phenomenon of anchoring with prior beliefs regarding an expected outcome. The authors gathered 70 undergraduates (30 male and 40 female) without any formal teaching training and assigned each of the participants to watch one of five conditions/videos where a child named Hannah performed her daily routine within different socioeconomic status situations (from high to low socioeconomic status). Some videos/conditions included indications of Hannah's academic performance.

The academic performance video showed Hannah answering correctly and incorrectly to easy and difficult questions making the child's performance appear inconsistent regarding her abilities. All participants after seeing the video(s) were asked to evaluate Hannah within four sections: (1) Hannah's academic level, (2) performance, (3) behavioural traits, (4) socioeconomic status. Participants were asked also whether they identified experimental conditions during the experiment, i.e., whether they noticed that some of the inputs that were provided were attempts to bias them.

Results showed that there was an expectancy confirmation effect on participants that viewed the academic performance video regarding the child's abilities demonstrated by a significant difference ($p < .001$), contrasting with the participants that did not view that video. Participants were also more inclined to report the child's socioeconomic background as high even without that kind of information, however, both groups (participants that visualised both videos and those who did not) yielded no differences regarding the use of socioeconomic status information to assess the child's academic abilities. The authors also observed that participants who were included in the expectancy confirmation conditions were more prone to confirm that the information provided about the child was "sufficient" to assess academic capabilities.

The confirmation bias phenomenon is intrinsically linked with the anchoring heuristic as well as with the representativeness heuristic, as it is easier to recall certain inputs that exist in greater frequencies within human reasoning (Klayman, 1995). It is also important to address the fact that selectivity of information regarding the type of evidence that individuals choose to support decisions seems to be unwittingly made, instead of being deliberately selected (Nickerson, 1998). That being said, people have a tendency to choose specific pieces of information and evidence, however, such a task will be carried out most likely without them having conscious awareness of that action (Nickerson, 1998).

A study by Bonefeld and Dickhauser (2018) in the field of education, asked 203 pre-service teachers to assess students from different ethnic backgrounds regarding their performance, based on two tasks (dictation and errors counted). Teachers received students' texts and their names which were either German or Turkish.

Results showed that there was a significant correlation between performance level, migrant background and implicit associations that teachers made ($r^2 = .573$, $p < .001$). On this correlation, the authors observed that teachers were more prone to claim that there were fewer errors in subjects with a German-sounding name even though the texts assessed were similar regarding the number of errors.

Another approach associated with the confirmation bias phenomenon is the association with previous answers or inputs from other people or entities since individuals, either lay people or experts, tend to confirm an initial hypothesis, rather than challenge them (Wason, 1960). This process, conscious or unconsciously made, is affected when information that threatens individuals' preconceptions is included in the reasoning process. Munro and Stansbury (2009) carried out a study where they focused on the process of challenging information, claiming that individuals have different attitudes within their decision-making process based on the level of challenge they experience regarding their preconception. Whereas people tend to disbelieve challenging information by presenting a sceptical mind, the opposite does not occur, i.e., non-challenging information promotes individuals to accept information at face value. In one of their experiments, the authors assigned 197 university undergraduates to groups of 10-12 individuals who completed experimental conditions which had two types of information, challenging and non-challenging information respectively, plus an individual assumption made by each participant. The level of threatening information was associated with a false (i.e., experimentally manipulated) prediction made by a fake expert. Results showed that participants who received challenging information that was counter to their first hypothesis were more prone to contest it than participants who received non-challenging information on the fake/manipulated prediction. This shows that challenging information regarding a previously made hypothesis by an individual can be interpreted as adverse whereas information that supposedly confirms the first hypothesis, even though erroneous, tends to be accepted.

The studies and data provided in this section have presented suggestions that were derived from experiments that used either trainees or students as participants. However, experts in different fields are also prone to unconsciously engage in a biased view of an event (Pines, 2006). Nickerson's (1998) work on confirmation bias provided a well-accepted definition of the term: "*the seeking or interpreting of*

evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand (…)" (p. 175). In this paper, Nickerson also provided a list of expert fields where one can observe decision-makers suffering from the confirmation bias. In medicine, for instance, a recent systematic review carried out by Saposnik, Redelmeier, Ruff, and Tobler (2016) analysed 114 publications with a total of 20 studies comprising 6810 physicians published between 1980 and 2015. In this review, the authors identified in every single study at least one cognitive bias or personality trait that affected physicians' performance. A total of 19 cognitive biases were identified to be impacting subjects' accuracy when diagnosing (36.5 to 77% of inaccuracy). Amongst the cognitive contaminations that were observed, there was in almost all studies the presence of the confirmation bias as one of the sources that affected physicians' performance. One of the main discussions in their paper was the fact that a large number of studies had raised the need to focus on the topic of cognitive contaminations, however, the authors still found similar results in recent studies, claiming that a shift is needed. The authors suggested the topic of cognitive biases should be included in medical training for established practitioners as well as for medical students.

A paper by Pines (2006) addressed the impact of the confirmation bias within the field of medicine. Pines described two scenarios where he recreated two potential realities regarding an emergency patient that was known in the emergency department to be recurrent and complaining about severe pain. In scenario 1, a nurse mentioned to the emergency physician that another emergency physician observed the patient and verified that everything was normal with his health. The physician accepted the information and confirmed results from previous standard tests and observations that were made in a crowded department. One can observe the tendency that the physician may have had to search for reasons aiming to confirm the information provided by the nurse and another physician on the night before. Scenario 2 presented a physician that tried to seek further information and challenge previous evaluations and comments from other peers. In the discussion, the author addressed the fact that professionals within the medical emergency field are "*susceptible to cognitive errors*" (p. 91) due to the fact of having a great cognitive load and being constantly challenged with knowledge requirements added to new situations they need to plan and manage. The author suggested that within scenario 1, there was greater space for confirmation bias to occur since professionals within

the medical emergency field will tend to search for specific information that gives them fast conclusions, as well as that, confirms either first hypothesis or peers' hypothesis and conclusions.

The search for confirmatory information promotes individuals to actively seek out pieces of information that confirm a hypothesis as well as to ignore relevant information (Klayman, 1995). A set of studies carried by Gilbey and Hill (2012) demonstrated this within the field of aviation. In this set of studies, a total of 170 individuals participated in their experiments, ranging from psychology students, student pilots, pilot orienteers and pilots with flying experience (between 20 to 160 hours). The authors asked participants to carry out three location strategy exercises for all groups in all experiments. In four out of five studies, the authors observed some participants (pilots, orienteers and psychology students) used a confirmatory strategy rather than a challenging strategy, i.e., these participants were prone to search strategies to confirm their hypothesis rather than trying to interpret information in a challenging way towards their first assumption. However, orienteers performed significantly better ($p < .001$) by using disconfirming strategies in some situations. Results also indicated that lay-people (the psychology students) performed poorly when compared with pilots or orienteers, confirming what the literature (Shanteau, 1988) within the "Experts' Decision Making" (section 1.3.) presented previously in this thesis.

Confirmation bias is a type of cognitive contamination that affects individuals' reasoning by tunnelling mental processes influencing one to ignore relevant information. The phenomenon is highly important to address as it happens with lay-people, but also with experts' decisions in specific fields which consequences are severe. There are more types of biases, and one of those that is intrinsically linked with the confirmation bias, as well as important for this literature review, is the phenomenon of having the context to be a source of potential errors, i.e., the information retrieved from the context where the decision is made and influenced by the context, named as contextual information bias. Both types of bias referred to are explored within the field of fingerprint analysis in section 1.7.3. as those are of interest to the empirical work conducted in this thesis.

1.4.2. Contextual bias

This thesis has been discussing the factors that affect individuals' judgements and until now, these factors were either related to people's level of expertise or to previous beliefs that influence one's decisions. However, there is another relevant factor which affects how people make their judgements, namely the contextual information that accompanies relevant information and its influence on individuals' decision-making processes.

In a paper by Todorović (2010), the effects of the context within the topic of visual perception were scrutinised. Throughout the paper, the author highlighted that the perception of a specific target is made not only by the visualization of the target per se but is also influenced by the effects that exist within the stimuli's surroundings, i.e., the information that individuals retrieve from the context in which the stimuli were being presented. Aronson (1999) also discussed the effects of the amount of information that individuals needed to make accurate decisions and the amount of information that one demands (i.e., the information that individuals mention they should have). The author claimed that having more information could be, in certain circumstances, helpful. However, too much information such as secondary non-relevant data could change the way the information is perceived.

A study carried out by Zukier (1982) asked 149 undergraduates to predict GPAs of target students. Participants received information correlated to the GPA of the target student which was highly or poorly correlated. In one of the studies, the experimental group also received non-diagnostic information (i.e. information that was not needed to make the decision and therefore would potentially dilute response accuracy) such as non-relevant tasks/activities the target student was enrolled in (e.g. visiting grandparents, number of siblings, hobbies). Although non-relevant information was not needed to assess the targeted students' GPA, the author found out that this type of information did affect participants' evaluation ($p\text{-}value < .05$), demonstrating that the dilution effect weakens focus that one gives to specific (and needed/required) information, promoting potentially erroneous judgements.

Research already stated that context such as appearance will affect how people tend to assess first impressions as Bateman and Mawby (2004) described. The authors elaborated four scenarios where they randomly assigned 307 respondents to

participate and face a manipulation of independent variables such as type of appearance of an interviewer that participants spoke with and the amount of information given by this interviewer, resulting in a composition of four-cell study design:

	Formal appearance	Casual appearance
Low information	Scenario 1	Scenario 2
High information	Scenario 3	Scenario 4

Table 3 – Bateman & Mawby (2004) study design

The authors observed the effect that independent variables had on the willingness of a participant to pay a value of money regarding the information presented. Results showed that there was an effect of the appearance of the interviewer as well as the amount of information provided:

	Formal appearance	Casual appearance
Low information	£13.66	£24.47
High information	£19.36	£32.29

Table 4 – Bateman & Mawby (2004) study results

A non-parametric (Mann-Whitney) test regarding the four scenarios was performed with some significant differences and some non-significant differences as well. The authors found that there was a significant difference due to the appearance effect ($p < .05$), however, there was not a significant difference regarding the information effect. When they analysed the combination of both effects, they only observed a significant difference between the extremes (formal + high information – casual + low information).

Veletsianos (2010) studied the effects that different types of contextual information associated with education agents such as teachers have on the learning process and observed this phenomenon at the early ages of education. The author assigned 94 students that were enrolled in elementary and childhood education technology courses in four experimental conditions regarding the agent (scientist or artist) and the type of tutorial (nanotechnology or punk rock). Participants were asked if they agreed that the educational agent was knowledgeable on the topic, only based on the appearance of the agent and by knowing what topic was to be taught.

Post-test questions asked all participants to recall all the elements they could remember from the pedagogical agent's presentation. Significant differences were found by the author regarding the 'educational agent' factor, the 'tutorial' factor and the interaction between 'agent' and 'tutorial' (*p-value* < .001), meaning that a large majority of participants based their answers on the contextual information they were provided with. Veletsianos (2010) also found that participants would recall more items from the artist agent rather than the scientist in both tutorial types (nanotechnology and punk rock), which may also suggest that appearance (i.e. contextual information) was a decisive factor for the learning process. Hence, the contextual information that was included in this experiment had an impact on the learning process despite the topic that was being taught.

Contextual information and its promotion of cognitive biases are extremely important, potentially impossible to avoid, and in some cases destructive as Clarke et al. (2000) presented in another study where they observed a 100% error rate in a sample of 97 trauma cases within an emergency setting. The authors mentioned that the largest set of errors involved contextual sources such as omission to consider relevant available information in the selection of appropriate care of the patients. The erroneous outcomes were not considered by emergency professionals as related to the reasoning processes they had, although these occurred in almost every case as observed by the authors, supporting what was previously mentioned by Tversky and Kahnemann (1974) regarding the non-logical reasons that individuals give to misguided decisions they make.

Egglin and Feinstein (1996) also observed the same type of error by randomly assigning six radiologists to two groups which were provided with a set of 24 pulmonary arteriograms with information regarding the prevalence of 33% of pulmonary emboli existence. To both groups, additional arteriograms information was also provided creating two conditions, one group with a final prevalence of 60% for pulmonary emboli and the other group with a prevalence of only 20% for pulmonary emboli. Participants were asked to estimate the prevalence of pulmonary emboli in each group and then cross over and review the groups' estimations after at least 8 weeks. Results demonstrated that diagnoses were significantly influenced by the context of interpretation as group A mean sensitivity for diagnosing

pulmonary emboli was significantly higher than what should be expected (75% versus 60%).

Even though contextual information is a potential source of bias, there is evidence that suggests other potential perspectives. A study by McRobert et al. (2013) found contextual information to have a positive influence on individuals' decisions. An emergency task environment with 9 skilled physicians (mean of experience = 20.1 years) and 9 less skilled physicians (mean of experience = 7.6 years) was simulated. Participants experienced two scenarios created to assess their performance. In both scenarios, there were two conditions of information (low and high context) that were provided to physicians during the experiment. Low context conditions only had contained relevant information whereas high context conditions also included non-relevant information. Participants had a set of points they would need to follow similar to a real emergency assessment for 20 minutes. In this study skilled participants seemed to be capable of dismissing irrelevant information considering it but not as relevant for diagnosis production as much as less skilled participants. Even when participants were not providing feedback to patients, they were asked to provide concurrent verbal reports (i.e. think aloud) during their assessment as Cormier, Pickett-Hauber and Whyte (2010) found that to be a good indicator of cognitive representations.

McRoberts, et al (2013) assessed their participants' performance by diagnosing accuracy, verbal reporting, and opinion-generation data. Results showed that less-skilled participants reported lower accuracy in their diagnosis (39%) than skilled participants (78%). Physicians with more years of experience performed all high-context tasks with no errors (100%) and more than half tasks correct on the low-context condition tasks (55%). Less skilled participants performed had a higher score on the low-context conditions (56%) whereas on the high-context condition they only had 22% correct responses. Finally, both groups had a higher accuracy score within the high-context condition. Concerning verbal reporting and data generation, skilled participants provided more inputs than their less-skilled counterparts performing better ($p < .001$) in some crucial tasks (e.g. evaluation, prediction, and deep planning).

The need to rapidly decide and having the context as a positive effect for the decision-making process was also demonstrated by another study carried by Rovira, Cross, Leitch, and Bonaceto (2013), where participants performed a rapid re-tasking exercise with the use of contextual information on their performance. Results suggested that contextual information was used to disambiguate and improve operator performance. Their experiment was conducted with seventeen cadets from a course in psychology at the U.S. Military Academy. The task was to choose from a setting with two sections [scheduler and the map] which of four options would be the most favourable to solve the task. In some parts of the Rovira et al. (2013) experiment, participants were given the same content of contextual information, although with different possible outcomes that they could achieve by considering the given feedback. The authors suggested that contextual information was a good variable to have in some of the training they conduct.

Individuals' decisions can be affected by different types of biases. This thesis has summarised two of those, the confirmation bias and the contextual bias. However, it is not always the case that biased reasonings will always lack accuracy, as the current section highlighted that the inverse is also possible. Regarding the possibility of having bias as a positive factor for accuracy, this thesis explores whether that also happens within the field of fingerprint comparisons. Also, the need to improve training is of relevance for the work this thesis is focusing on and this is discussed within the overall discussion (Chapter 6). When professionals are aware of the potential pitfalls originating from sources of errors (such as cognitive biases) they may become better prepared to achieve positive outcomes, however, they need to be trained to recognise these potential errors.

1.4.3. Debiasing strategies

Even though cognitive contaminations seem to be impossible to avoid, either by individuals who achieved a high level of expertise in their knowledge within a certain field or by lay-people, the fact is that one should not assume that decisions will always be biased, i.e. erroneous, and therefore not consider possible changes and improvements. A possibility that has been used in social science disciplines, but not limited to those (e.g. physics) is the blind analysis approach (MacCoun & Perlmutter, 2015). According to Roodman (2003), the technique itself is meant to

measure something without looking at the answer. This type of technique is an "*optimal way to reduce or eliminate experimenter's bias and the unintended biasing of a result in a particular direction*" (p. 1). There are different types of ways to apply a blind analysis. One can simply scramble labels of data, ask other colleagues to assess the same information without knowing what they are assessing or even arrange a procedure where the same person assess the same data without knowing which data, specifically, he/she is analysing (MacCoun & Perlmutter, 2015).

Montibeller and Winterfeldt (2015) suggested different types of debiasing strategies that individuals may use to mitigate the consequences of different biases. Some of the strategies that the authors suggested focused on preventing cognitive biases such as anchoring or confirmation biases, which were claimed to be harder to mitigate. For these types of biases, they suggested techniques that asked individuals to (1) separate important factors within the tasks (e.g. separate value and utility elicitation), (2) use teams with multiple assessors that do not know previous answers, (3) conduct statistics training and (4) use fixed values instead of probability elicitations. The authors also mentioned that confirmation biases are related to emotional cues, i.e. that have an influence related to the emotional response that unconsciously individuals provide for a certain problem/task. Regarding those types of biases, they provided some strategies to mitigate them, such as (1) avoiding loaded descriptions of consequences in the attributes of the problem, (2) use multiple experts to carry out the same task and explore alternative points of view or (3) challenge the outcome that previously was achieved.

With a focus on design thinking, Liedtka (2015) also proposed strategies comparable to the previous ones. The author suggested that biases could be related to flaws that decision-makers have when they hypothesized testing strategies. These flaws were associated with (1) an excess of optimism, (2) the inability to challenge data, (3) an attachment to early solutions/conclusions and (4) the preference for options that could be easily imagined (correspondingly to a heuristics process). It was also suggested that decision-makers work within an environment with multiple opinions, and seeking to actively disconfirm data, i.e. to work in contexts where scientific reasoning could be applied as well as hypothesis testing.

MacCoun and Perlmutter (2015) also suggested five techniques – noising, biasing, cell scrambling, item scrambling, and various combinations – to mitigate confirmation biases. Those techniques were all implicitly linked with the action to implement a blind analysis strategy as described in the table below.

Technique	Perturbation	Potential application
Noising	Add a random number (from an appropriate statistical distribution) to data points or model parameters.	Testing which of several prevention messages is most effective in reducing smoking.
Biasing	Obscure differences in experimental conditions by adding a hidden value that is biased in a particular direction.	Estimating whether the cost of a controversial safety regulation exceeds its benefits.
Cell scrambling	Shuffle labels for experimental conditions, so that it is unclear which set of results matches which conditions.	Testing a prediction that hard-copy books are better comprehended than audiobooks.
Item scrambling	Randomly relabel each data point to de-identify experimental conditions.	Analysing group differences that might be easy to recognize even with noise and bias (e.g. effects of neighbourhood and school on crime victimisation).
Various combinations	Row scrambling [keep pairs of variables together to preserve correlation] Variable blinding [swap labels of various variables]	

Table 5 - MacCoun and Perlmutter (2015) strategies to mitigate confirmation biases

Another strategy that was suggested by Croskerry, Singhal and Mamede (2013) was the use of feedback. However, the use of feedback will be scrutinised in the following section focusing on the importance of motivation for decision-makers.

Throughout this section, different ways that decision-makers can be influenced in their decisions and commit erroneous decisions were explored. As seen, some of those influences may pass unnoticed and without conscious acknowledgement by individuals. One of the most important points from this section is the fact that people need to become more aware of biases and implement strategies to prevent

them in order to avoid misguided decisions. Another important note from this section is the fact that a biased decision can be committed by both lay-people and experts. Even though experts show higher accuracies, they had their decisions biased in the studies that were analysed. It is also interesting to note that the majority of the strategies that research has been suggesting for bias mitigation, involve sharing the tasks amongst more people and challenging hypotheses. These strategies have been applied to fields such as psychology and sociology in the past, and recently they have been implemented in STEM fields and even more recently, they have also been suggested by official reports to be implemented in the central field of this thesis, forensics. Other strategies to mitigate biases such as feedback are covered in the following section where this thesis describes a final important variable for the process of decision making, namely motivation.

1.5. MOTIVATION AND DECISION MAKING

Whether expert or non-expert, when one aims to explore decision-making processes and the variables that affect performance, either positively or negatively, motivation – as a concept – should be of interest. As in other concepts, there are several ways to define motivation, even though almost all are originated from the Latin etymon *movere*. A broader definition of motivation can be found in Kanfer's (1990) suggestion where the author presents motivation to be the "*psychological forces that determine the direction of a person's level of effort, and a person's level of persistence in the face of obstacles*" (p. 80). A more specific and rather useful definition of motivation that was adopted in this thesis is the definition suggested by Hoffman (2015, p. 8) that stated motivation to be "*the degree of effort and intensity toward a goal related to learning or performance*".

Motivation is a core factor for decision-making processes as its exploration allows and facilitates the understanding of individuals' behaviour; it is a combination of needs, drives, interests, attributions and intentions (Jones & George, 2008). According to Tversky and Kahneman (1974), motivation plays a key role in how individuals make their decisions since they use different strategies, which may change due to experience, context, goals, and certainly, emotional drives. Even though there is diversity within the categories that one can analyse when assessing motivation such as (1) academic, (2) achievement, (3) biological/evolutionary, (4)

choice, (5) optimal, (6) performance, (7) personal and (8) social (Hoffman, 2015), researchers should be aware of two basic distinctions of motivation, *extrinsic* and *intrinsic* (Deci & Ryan, 2000). Whereas extrinsic motivation influence individuals' behaviour and attitudes due to external sources, (e.g. social rewards or punishments), intrinsically motivated people are motivated by doing certain actions, and those actions motivate them to keep doing them as if it was a cycle (Jones & George, 2008).

The differentiation of the types of motivation that people can demonstrate during their work was also described concerning the behaviours exhibited when carrying out a task or learning a new thing. According to Hoffman (2015), intrinsically motivated individuals have their personal goals maintained by intrinsic self-improvement, i.e., by the sole task of learning and being cognitively challenged. These individuals have what Hoffman described a 'Mastery Approach Orientation' as they "*seek knowledge under the presumption of personal improvement*" (p. 159) and have the desire to make decisions that maintain their psychological equilibrium. On the other part of the spectrum, Hoffman (2015) described the 'Normative Approach Oriented' individuals. These were illustrated by individuals who would achieve tasks due to the pressure of peers (or the social environment) and would benefit from other sources of motivation besides intrinsic ones, thus would be also influenced by other sources of stimuli than what the task provides them. Regarding biased decisions related to the types of motivation, Hoffman (2015) claimed that erroneous conclusions can usually be observed in individuals who are influenced by external sources.

There are different means to assess motivation as there are different ways to conceptualise it. For this thesis, it is important to focus on the way to analyse the different aspects of motivation that Hoffman (2015) described, the intrinsic and extrinsic motivation that individuals present in order to differentiate mastery oriented individuals from normative oriented individuals as the later can be more susceptible to be influenced by external sources such as cognitive bias.

Tremblay, Blanchard, Taylor, Pelletier and Villeneuve (2009) explored those aspects and designed the Work Extrinsic and Intrinsic Motivation Scale (WEIMS). This psychometric tool was designed based on the work that Deci and Ryan (2000)

conducted within self-determination theory. It has been used to assess individuals in different domains such as Occupational Therapy (Chai, Teoh, Razaob & Kadar, 2017), Management (Nordhall & Knez, 2018; Shu, 2015) or Social Work (Proença & Cristina, 2013) and was reported by Hoffman (2015) as having internal consistency to analyse both aspects of motivation. However, there is no rigid way to divide people's motivation, some people are more intrinsically motivated, some others extrinsically, and there are also people being motivated by both types. Jones and George (2008) illustrate the mix of having both types of motivation within nursing where people "*enjoy taking care of patients*" but also because he/she "*has a secure job with good benefits*" (p. 521). The types of motivation that one has, depend on three different factors according to Jones and George (2008): (1) personal characteristics (e.g. values, attitudes, personality and abilities), (2) the nature of the jobs and (3) the nature of the organisation and its features such as culture and structure, human resources management, control system, payments and other organisation of rewards.

Cooper (2002) defined the existence of two main types of factors which influence individuals' motivation: (1) Personal factors and (2) Situational factors. Personal factors are related to features associated with psychological qualities such as disposition, temperament and intelligence that promote individuals' abilities and specific skills. Situational factors are focused on variables the person may not control entirely, such as quality control systems, the size of the organization where he/she works, the type of co-workers, type of communication, norms, reward systems and management practices. Both factors are closely linked, and together they form part of what is analysed when looking at one's performance such as effectiveness, responsibility, autonomy, confidence, creativity and satisfaction.

Due to their interests, when individuals are challenged with certain types of tasks, they may develop a certain level of expertise associated with work satisfaction (Ackerman & Beier, 2006). This satisfaction can influence individuals' confidence to carry out a certain task (self-efficacy) as well as the competence estimation that the individual has when they perform a specific task (self-competence). Although individuals must acknowledge what can they do, and fairly estimate their confidence in doing something. Holland (1997) suggested that people's drive to

excellence needs to be fulfilled by a motivational factor, even in tasks they already have experience.

Having that in consideration, it is very important to analyse individuals' motivation and the association with their level of expertise or skills acquisition. Within this thesis, the term and value of motivation are extremely important due to the fact that participants in the experimental tasks were assessed not only on their level of expertise and the potential comparison that can be done with lay-people. In addition to the potential biases that respondents of the experimental tasks were exposed to, they were also evaluated on their level and type of work motivation. Therefore, motivation played a crucial role to better understand some of the characteristics that fingerprint examiners have personally and professionally.

1.5.1. Motivation, Performance and Need for Cognition

It is possible to observe decisions that people make without being consciously effortful, whereas some other decisions require a greater cognitive effort. Verplanken and Svenson (1997) elaborated three different types of decisions people make daily: (1) High-cost decisions (e.g. buying a house), (2) far-reaching consequences decisions (e.g. choosing a career) and (3) significant opinion or emotional value decisions (e.g. voting for a certain political party).

Even though some studies ignore motivation, there has been some research that highlighted its importance such as the work carried out by Kardes, Muthukrishnan and Pashkevich (2005), who stated that motivation has an important role within the process of making a decision. These authors reviewed a number of studies and challenged the idea that judgemental bias could decrease as experience and motivation increased. Their review claimed that individuals who are more motivated to carry out a certain task, and who also have more experience do not need to be exclusively biased.

Regarding motivation and its implications on decision making, Kardes, Muthukrishnan and Pashkevich (2005) claimed that individuals can have their judgments flawed due to an excess of interpretation processing, i.e. when individuals process a large amount of irrelevant information that is a source of

motivation, people tend to overinterpret and overuse information that is not as important as it seemed at first sight. In other words, people could become biased because they are motivated by the availability of the information they did not need to make a decision.

Regarding the claim made by Verplanken and Svenson (1997) that pointed out the lack of literature between motivation and decision-making processes, Mantel and Kardes (1999) proposed a model to explore the effects of motivation on making judgements. They proposed that motivation was determined by two variables, (1) need for cognition and (2) involvement.

Need for cognition was described as the effect that people experience when they feel cognitively challenged by performing certain types of tasks and experience a greater level of enjoyment carrying out those tasks (Cacioppo, Petty & Kao, 1984). The construct was first developed from the work of Cohen, Stotland and Wolf (1955) who described the need to attend and analyse the phenomenon they observed in people who enjoyed carrying out a task just for the challenge the task provided them. Following the discussion of Cohen, Stotland and Wolf work, Cacioppo, Petty and Kao (1984) designed the Need for Cognition Scale (NCS) in order to evaluate the level of enjoyment one has when he/she is engaged in certain tasks. There will be different ways to engage within the same task by different people. For that reason, further works described that people with low levels of Need for Cognition usually utilise shortcuts such as heuristics to conduct their decision-making processes (Coelho, Hanel & Wolf, 2018). When people perform tasks which they enjoy and where the cognitive effort is high, individuals will internally search for higher cognitive skills to perform the task better (Mantel & Kardes, 1999).

The second variable described in Mantel and Kardes model, involvement, is related to the decision-makers' experience regarding the context in which the decision happens, i.e. the motivational conditions that are driving the decision, either positively or negatively. High involvements are associated with a greater cognitive effort since people tend to become more dedicated and exert greater effort when they ascribe higher importance to something (Mantel & Kardes, 1999).

There are three types of involvement according to Verplanken and Svenson (1997). The first type is the *value relevant of involvement* which is related to the psychological state linked to important values associated with human attitudes. In this type of involvement, it is possible to find three positions that people can take. Either they (1) accept, (2) they reject or (3) they do not commit. The second type of involvement is *impression-relevant involvement*. This kind of involvement is described as the extent that one's self impression is expected to affect others' opinions. Finally, the third type of involvement, *outcome-relevant involvement* refers to the weight of the arguments that are used within a decision-making process. Strong arguments will affect one's own position and on the other hand, weak arguments will be less effective in people's position when it comes to making a decision (Verplanken & Svenson, 1997).

1.5.2. Feedback in Decision-Making

The action of asking for feedback should be seen as of important value. Not only was it acknowledged as a potential strategy for mitigating risks of bias (Croskerry, Singhal and Mamede, 2013), it is also a way to achieve performance improvements. According to Boud and Molloy (2013) feedback is "*a process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standards for any given work, and the qualities of the work itself, in order to generate improved work*" (p. 205).

In a study on Parkinson's disease, Osman (2012) observed the importance of feedback between a practitioner's and a patient's perspective. According to the author, feedback can guarantee an order to organise future actions and also to mediate and evaluate some of the ongoing tasks (Osman, 2012) with higher accuracy for medical staff and higher receptivity for patients. Feedback can be seen as part of the sequential model that decision-makers need to pass through (Brehmer, 1992) as it was noted in section 1.2., people *prime* and *perceive*, and then they *act* in a certain way which may provide some learning from the experience. Hence, feedback needs to be considered as part of this decision making sequence.

Campbell and Waters (2013) claimed the need to have an analysis of feedback from medical professionals aiming to improve the methodologies within an urgent setting

of treatment. This analysis of feedback facilitated the development of innovative methodologies, such as algorithms, which can be used to produce better procedures that professionals follow in their day-to-day practices.

The work of Jessup, Bishara and Busemeyer (2008) also demonstrated the importance of feedback. Participants in their study were divided into two groups, one received feedback during the task, whereas the other did not receive any feedback. The authors mimicked a gambling task where participants needed to decide how much risk they were willing to take, with a guarantee that they would win a small but certain amount of money, or if they wanted to take a more substantial risk, without any guarantee, to win a larger amount. Results demonstrated that the existence of feedback within the procedure changed the way decision-makers perceived the information and the weight they attributed to each option. Participants who received feedback tended to underweight the small probabilities and chose the smaller, but certain, amount of money to earn. Individuals with no feedback had the opposite result, they preferred to choose the bigger and uncertain amount of money and over-weighted the small probability associated with winning.

The work of Bahrami et al. (2012) observed the effects of having feedback and the role of social interaction among participants during the tasks they completed. The authors constructed three different experimental settings where 72 adults were assigned (in pairs) to one of the 3 conditions. The task consisted of answering questions that were presented in a computer-based experiment. In scenario 1 and 2 participants could communicate when a disagreement occurred, in scenario 3 participants did not have an opportunity to negotiate. In addition to the fact that private answers were given before any negotiation, participants in scenario 1 and 3 were also given feedback (correct or incorrect) related to their answers. Results suggested that feedback influences individuals' accuracy by guiding them to what should be the best practice to follow. However, the social interaction among participants provided benefits such as building up reliable and collaborative tasks and enabling the participants to achieve an accurate result within the experimental task.

Sterman (1989) claimed that, although feedback did have an important influence on the decision-making process, sometimes feedback itself is misinterpreted by people leading to poor performance. The experimental design lasted for four years and had 192 individuals participating. The task that participants in this study needed to do was based on playing the 'Beer Distribution Game' (Supply Chain Academy, n.d.) developed by MIT which consisted of having four sectors (1 – retailer, 2 – wholesaler, 3 – distributor and 4 – factory), with each participant assigned to a sector. Each participant's objective was to spend the optimal amount of money regarding their sector's responsibility (e.g. a retailer that needed to do an optimal investment on beer cases to sell all the beer). In order to achieve optimal performance, individuals could ask questions to gain knowledge and feedback about their ongoing decisions. Results showed that several players had their performances lower than the expected due to irrelevant feedback, caused by participants searching for irrelevant information.

Feedback should not be perceived only as a top-down task. It impacts both sides, the one that provides feedback, as well as the person that receives it. Boud and Molloy (2013) claimed that the process of providing feedback does not need to be exclusively a "teacher-student" relationship. The authors suggested that this process can become richer by retrieving data from third parties that can contribute to the learning experience. Boud and Molloy (2013) also suggested that people who manage the process of providing feedback should analyse the purpose of providing feedback as well as when is it allocated and in which format is it provided. In other words, these authors claimed that feedback should not be only a simple 1:1 conversation, but rather a process with rules and guidelines, aiming to promote individuals' performance at different stages of their learning process.

According to Kahneman (2011a) feedback, when regularly provided, is a strategy that increases performance when individuals need to make decisions that have high consequences. The author claimed that the sooner feedback is received, the greater impact it will have on future actions. Tsai et al. (2015) after looking at feedback systems in health environments associated with clinical decisions argued that the analysis of feedback in some environments should be interpreted in order to develop accurate guidelines for further decisions.

From this section, it is important to note two things. Firstly, it seems clear that motivation plays a role in the process of making a decision, and that people have different types of motivation. There is always a role of motivation, and it will define how the individual perceives the decision he/she needs to make. Considering that people have different types of motivation, it is expected that they will also approach the same task differently sometimes. Difficult and easy decisions are carried out in different ways, however, the involvement one has will define the commitment and potential influence that external sources may have on the outcome. It was explained throughout the section that motivation plays a specific role in the involvement one has, and therefore the less the person is involved, more influenced he/she can be, by cognitive biases. A strategy to improve performance is to work on the motivation individuals have and the value of reflecting on their own actions.

Feedback should not be exclusive to a one to one relationship. It can include data from third parties when needed and feasible. However, it needs to be done under certain conditions to avoid confusion with simple 1:1 conversations. The goal of providing feedback can be the improvement of one's performance, and to achieve that, it should be carried out in a standardised way.

1.6. DECISION MAKING AND COGNITIVE BIAS IN FORENSIC SCIENCES

Previous sections explored the general literature related to decision making and related concepts that can influence individuals' decisions such as cognitive bias or motivation. It was identified that those concepts are of great importance to understand how experts differ from laypeople, as well as to enhance individuals' performance. In this section, this thesis funnels the literature reviewed previously into the population of interest for the work conducted, namely fingerprint examiners and their field of work, the forensic sciences. As it will be observed, the concepts related to the decision-making processes and the effects that cognitive bias has within it have been of particular interest to the forensic community since it allows the identification of gaps in knowledge and practice.

In the forensic sciences, the impact of bias is of great interest. One of the most important aspects of this impact is the notion of justice within society. Criminal

Justice Systems that do not seek to minimise flawed decision making within evidence examination, may challenge justice and its fairness. This can be observed by the occurrence of cases where miscarriages of justice have been verified. Between 1989 and 2017 the National Registry of Exonerations (NRE) in the U.S. identified 2,169 exoneration cases (NRE, nd). The same entity has been providing reports, the most recent released in 2018, which analyse contributing factors to wrongful convictions. Wrongful convictions reported by the NRE had their origin either in procedures of the criminal investigation (e.g. mistaken eyewitness identification or false confessions) or in forensic evidence (e.g. DNA evidence).

In total, in 2018, the NRE observed that wrongful convictions totalled over 1,639 years in prison, with an average of 10.9 years per exoneree (NRE, nd). Issues such as improving methodologies that forensic scientists use, gaps for potential errors, and the gaps within the fields of forensic sciences where biases can be observed are the main issues that this section discusses.

Aligned to the NRE reports, it is possible to find a number of similar cases in two legal projects, the Innocence Project [USA] (Innocence Project, 2019) and the Innocence Network [UK] (Innocence Network, 2019). The Innocence Project reported (Innocence Project, 2019) 116 cases which were part of a larger group [*n* = 225 cases] that had several procedures, within forensic disciplines, that were questioned regarding the methodologies used. In these cases, more than 50% involved the application of improper procedures and the Innocence Project has since been involved in high profile cases such as the case of Steven Avery who was wrongfully convicted of rape, spending eighteen years in prison. Furthermore, the Innocence Project has also passed 15 wrongful convictions reforms[1]. Most of the cases include contributing factors associated with procedures or erroneous testimonies. The evidence that is most used to exonerate cases is DNA, however, cases that have been sentenced had other types of evidence such as fingerprints, eye-witnesses, fibres, and others. Even though this area of challenging forensic evidence and exonerating cases that were wrongfully convicted is extremely important, the Innocence Network in the UK has stopped its activity since 2014. On

[1] The numbers presented in this thesis were verified on the 26th June 2019. Access at https://www.innocenceproject.org/justice-2018/

the Innocence Network UK website, one can read that the project itself successfully achieved some of the aims it was proposed to, specifically to disseminate the need to investigate potential wrongful convictions. Additionally, the website also mentions some issues that limited the sustainability of the project, amongst them, the need for funding to continue the *pro-bono* work (Innocence Network, 2019).

Since forensic sciences involve decisions made by human experts, one can expect to observe similar issues to those described in section 1.4. (i.e. cognitive bias such as confirmation and contextual bias based on heuristics that affect experts). The state of the art within the field of decision making in forensic science gained attention since media covered specific cases such as the Mayfield case (Gavett, 2012). Nevertheless, it is possible to observe in the literature some factors that have gained attention from research (e.g. contextual bias or human factors).

Laypeople may expect that forensic science and the criminal justice system are immune to errors. However, miscarriages of justice as reported by the NRE or the Innocence Project, suggest that forensic science is vulnerable to errors. Only using three key-words (forensic + science + errors), a web-browser search such as Google, provides more than 9,000,000 hits in English, +6,000,000 hits in Spanish and Portuguese almost 500,000 hits.

Since there are a considerable number of variables, some of them associated with technology and some of them associated with human performance, the potential for errors is still high due to human expertise being of the primary diagnostic tools used during forensic examinations (Dror & Cole, 2010). A study carried out by Kassin et al. (2013) observed the phenomenon of confirmation bias within the forensic domain. The authors referred to it as "*forensic confirmation bias*" (pp. 45). Kassin and colleagues explained this new concept as the effects that forensic examiners experience due to the influence of their own beliefs, expectations, perceptions and motivations as well as the contextual information examiners are often exposed to.

There are numerous fields within forensic sciences which are at risk of the effects of cognitive contaminations. Some erroneous decisions are made due to wrongful expectations (Kassin, Dror & Kukucka, 2013) as well as to the so called *elasticity* of the forensic evidence (Ask, Rebelius & Granhag, 2008) which was scrutinised in

one study that assigned 117 police trainees in Sweden to 1 out of 6 possible conditions defined by 3 types of forensic evidence (DNA, photo and witness) versus 2 types of evidence outcome (consistent and inconsistent).

The case that participants were invited to take part in was an illustration of a homicide case which promoted high involvement and motivation from participants. Evidence was given in a fixed sequence. First, the DNA evidence (analysis of a bloodstain in the suspect's jacket), followed by photo evidence (pictures from a taxi camera) and thirdly the witness testimony. Evidence was also consistent or inconsistent with the identity of the suspect. Participants were first asked to provide their insights using a scale of 1 to 9 regarding the probability of the suspect having committed the murder (1 – not at all probable to 9 – absolutely certain) and regarding the strength of each piece of evidence presented (1 – very weak to 9 – very reliable). Then, participants were provided with another booklet with evidence that either was consistent or inconsistent with the association of the suspect to the crime and asked again to provide their insight regarding the guilt of the suspect using the same scale. Finally, respondents were solicited to address which reason was the most important to support their answers (open-ended question).

Results showed that participants in the consistent evidence condition attributed greater reliability to the DNA and photo evidence ($p < .001$) after receiving more information, suggesting the occurrence of confirmation bias in their assessment process once that in their first answer they rated a guilt expectation valued of 7 points, and after consistent evidence being received they changed their answer for an upper value in the guilt expectation scale. Elasticity was observed and demonstrated by the results on the witness evidence variable which, depending on its consistency, was rated significantly different regarding the type of consistency of evidence they received (*M-consistent* = 6.69, SD = 0.87 vs. *M-inconsistent* = 5.10, SD = 1.09). In the arguments that participants were asked to provide at the end of the experiment, the two most important arguments for photo and DNA evidence were related to being made by forensic expert's reports.

Osborne, Wood, Kieser and Zajac (2014) also explored the influence of cognitive contaminations on the decision-making processes of forensic scientists, specifically regarding the effects of contextual information within bitemarks comparisons. The

authors asked 178 participants that were enrolled in different dental studies (dental surgery, dental technology and oral health) to analyse 96 pairs of bitemark impressions and dental overlays. The total set was then split into two sets of 48 pairs. Regarding the information that each pair had, each of the sets included ambiguous (non-relevant) and unambiguous (relevant) information, i.e. information that lacked quality and information that was fundamental to make a decision with quality. To observe the influences of contextual information, the authors added pictures of violent crimes to 1/3 of the pairs. They also added to another third of the set a priming wording (i.e. words such as "same" and "guilty") that were meant to promote a confirmatory bias in the participants' responses. Results showed that within the trials where there was no context included, the accuracy of matches increased, contrary to what happened in the trials where non-relevant context was provided to participants ($p < .05$). Regarding the academic training that participants had, differences within the studies' results were also found. Respondents that were enrolled in dental studies had higher rates of matches than non-dental students ($p < .05$), showing an effect of what the authors suggested to be an association with an expert level of knowledge since dental students had learned about bitemark analysis.

Regarding the potential ambiguity in forensic evidence, a study focusing on DNA evidence conducted by Dror and Hampikian (2011) demonstrated that examiners may reach different conclusions when analysing genetic material. Authors gave the same DNA evidence from a rape case to two different groups of qualified DNA expert analysts. Group 1 analysed the DNA evidence without any kind of extraneous contextual information whereas Group 2 received contextual information regarding the guilt of one of the suspects when they were analysing the evidence. Results showed two important aspects regarding the variability of DNA analysis. First, the authors mentioned that analyses were not entirely objective, which was not to be expected as DNA used to be considered the *gold standard* of forensic science. Secondly, it seemed that examiners had their outcomes biased by the extraneous contextual information as only 1 out of 17 examiners in Group 2 reached the same conclusion as examiners in Group 1.

Invalid forensic procedures and wrongful convictions associated with forensic evidence, has brought attention to the issue of bias mentioning a wide range of fields

such as fingerprints, bitemarks, bloodstain, DNA, handwriting, hair analysis, amongst several others (Cooper, 2019) where erroneous decisions were found in cases of post-conviction DNA exonerations. After analysing 137 trials transcripts, Garret and Neufeld (2009) found evidence of invalid testimonies by forensic practitioners in 82 cases (60% of total cases). The sample of invalid testimonies comprised 72 forensic examiners from 52 laboratories, practices or hospitals in 25 different states in the U.S. During this study, different case trials where errors were presented had different sources, such as invalid scientific procedures, which led judges and jurors to reach incorrect conclusions.

Considering past miscarriages of justice, some official reports have been released for the public, with a special focus on the forensic science community. Among these, are three important reports, the National Academy of Science's Report "Strengthening Forensic Science in the United States: A Path Forward" (NAS, 2009); the Report to the President – PCAST Report (Executive Office of the President, 2016); and the Forensic Science Regulator Series (FSR, 2020; FSR, 2013). The NAS report made a number of key recommendations such as the need to have stronger ties between forensic laboratories and universities in order to carry out research and improve procedures, as well as having forensic professionals undertake updated training programmes (NAS, 2009). Another very important recommendation the report makes is linked to the resources (technical and human resources) the laboratories/offices currently have. This point is indeed important since a lack of human resources might be one of the main reasons contributing to variables such as time pressure affecting practitioners' performance negatively (Huber & Kunz, 2007; Maule & Edland, 1997).

The NAS report presented three main challenges related to the lack of resources within forensic laboratories: (1) how to inform the investigation teams effectively, (2) how forensic scientists can provide the prosecution with reliable evidence and (3) how can forensic scientists avoid errors that lead to miscarriages of justice (NAS, 2009). Throughout this report, there is a diversity of reference to the need to have forensic practitioners work closely with institutions such as police forces or national justice agencies. However, the report also clearly stated the necessity to promote the independence of forensic bureaus from police agencies, aiming to prevent errors within decision making due to pressures or other types of biases.

To meet the challenges that forensic bureaus have been facing, the National Institute of Forensic Science was created. This institution was required to include in its structure a strong culture of research, as well as a teaching culture including federal laboratories and smaller laboratories such as state forensic agencies (NAS, 2009). The release of the NAS report promoted a number of changes in a variety of fields within forensics, including fingerprints. Recently with the election of President Biden in the U.S., the discussion has been put forward again aiming to create what the NAS report suggested when published – the creation of an independent entity to oversight the work within the field of forensics (Smith, 2021)

Another report, also produced in the U.S. – Report to the President, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, also known as the PCAST Report – is one of the most recent reports in this field (PCAST report – Executive Office of the President, 2016). This report aimed to strengthen the scientific underpinnings of the forensic disciplines and indicated a variety of suggestions for three organisations: The National Institute of Standards and Technology, the Office of Science and Technology Policy and the Federal Bureau of Investigation Laboratory. The PCAST Report also made a set of recommendations for the Attorney General and the Judiciary in order to promote the more rigorous use of forensic sciences within the criminal justice system. Within the report, a number of forensic disciplines (e.g. DNA, Bite-marks, Latent marks, Firearms, Footwear and Hair analysis) had their validity analysed as well as their procedures. As mentioned above, one of the most important aspects of the PCAST report was the validation and standardisation in the field of forensic science.

Within the PCAST report, there were sections dedicated to the criteria for validity and reliability of forensic sciences' comparison methods, where issues such as cognitive bias or proficiency testing have their influence. Within the cognitive bias section, the report claimed that human perception can be distorted and thus cause misguided judgements shaped by factors such as contextual information, (i.e., judgements that are influenced by irrelevant information), or by confirmation bias, where forensic experts anchor to new information, which may be irrelevant and may make these professionals conform with pre-existing beliefs (Executive Office

of the President, 2016). A suggestion the report makes within the topic of cognitive bias is based on methodologies in the biomedical community:

> "*The biomedical science community, for example, goes to great lengths to minimize cognitive bias by employing strict protocols, such as double-blinding in clinical trials.*" (p. 31)

The other section of relevance for this subject within the PCAST report focuses on proficiency testing. Possible solutions the PCAST report has suggested to prevent errors caused by contextual information within laboratories is the implementation of proficiency testing. Proficiency testing has been widely used in different fields and the report suggested its application to the wide range of forensic disciplines. As in other fields, proficiency testing is claimed to be the "*the only way to establish scientifically that an examiner is capable of applying a foundationally valid method is through appropriate empirical testing to measure how often the examiner gets the correct answer (…) often referred to as proficiency testing.*" (p. 57). The report is very specific on the definition of proficiency testing since the term is sometimes wrongly used either to describe if examiners follow procedures without having their accuracy assessed, or to describe exercises that promote technical improvement overcommitted errors without post-analysis on the casework that is carried. Hence, PCAST report states proficiency testing as:

> "*ongoing empirical tests to evaluate the capability and performance of analysts.*" (p. 57)

One should also note that the report distinguishes proficiency testing from competency testing. The latter is considered to be the individual's assessment of capabilities prior to performing a certain set of tasks (ASCLD/LAB, 2011). Finally, the report added a very specific feature on how to apply a proficiency test, within a blind context. This type of implementation prevents forensic practitioners suffering from what has been named as the *Hawthorne Effect*, where individuals have their attitude towards a certain task or set of tasks affected by the fact that they know there is an assessment being conducted. However, the report also mentioned the fact that it is not possible in every forensic laboratory to implement blind

proficiency testing. It does seem that this type of implementation depends on the structure of the laboratory as well as the type of forensic discipline.

This thesis explores (Chapter 4) the differences between approaches within different laboratories, which may affect the implementation of this type of testing within the field of fingerprints, namely regarding the accreditation and/or technical approach that examiners need to follow within their casework. The National Commission on Forensic Science (NIST, 2016) stated some guidelines regarding accreditation within forensic laboratories, where blind proficiency testing was suggested as a valid strategy to comply with accreditation. In this document, the abovementioned entity referred to blind proficiency testing to be an approach where the individuals who work there do not know they are being tested. In the field of DNA, Peterson, et al (2003) tested the feasibility of having external blind proficiency testing in U.S. forensic DNA laboratories. In their study, the authors observed a difference of time in the turnaround that public and private laboratories had. Public laboratories were slower in their turnaround than private laboratories that work independently and in a commercial setting. This observation is quite important. On one hand, forensic science has received several guidelines regarding the need to implement strategies to comply with accreditation. However, on the other hand, resources in the public domain cannot equal the performance of private laboratories.

One can think about the scarce resources that the public laboratories have when compared with private ones (Gallop, 2020). This difference implies a limitation to implementing blind proficiency testing to a high standard. Another limitation regarding the implementation of blind proficiency testing is the type of implementation and delivery. The research argues that blind proficiency tests should be implemented externally (Saks, 2005). However, funding availability in forensics has become a problem to solve as well (Gibb, 2019), which amongst the lack of resources such as materials and human resources, the investment to outsource blind proficiency testing is a significant barrier to implementation.

In the United Kingdom, the Forensic Science Regulator (FSR) fulfils a similar role to the American institutions cited previously. The FSR is the entity in the UK that provides guidelines for quality standards and good practices within forensic

services (FSR, 2013), however, the FSR has no statutory powers in the UK, which limits their ability to enforce guidelines.

One of the responsibilities the FSR has, is the responsibility to require quality standards for new or improved techniques, to lead the development of new standards, and to provide advice and guidance for forensic services in order to make them work within the required standards before being allowed to be admissible in court. To promote new standards, the FSR has published guidelines for good practice for forensic disciplines such as DNA and fingerprints analysis as well as a consultation document focusing on cognitive bias in forensic sciences (FSR, 2015).

These guidelines are in agreement with the American PCAST report (2016) that advocates the management of the information flow using a linear sequence and advising analysts to prevent access to non-relevant information by forensic examiners. However, the PCAST report stated a further step that the NAS report did not, the introduction of blind proficiency testing.

In the Forensic Science Regulator's guidelines (2015) one can observe a compromise between the two American documents, where examiners should have within their case-work a blind reviewer, i.e. that examiners who conduct forensic analysis should have another examiner that assesses all the work within a blind setting. Regarding proficiency testing, the FSR document suggested laboratories should implement proficiency testing, but there is no mention of this type of quality measure using blind testing.

Besides national reports, the academic research community has been working towards the improvement of forensic professionals' performance. Kassin et al. (2013) suggested guidelines for practitioners to follow. Firstly, they suggested the need to have a workflow based on a linear sequence, i.e. the workflow should be sequential and direct, instead of being carried within a circular setting where potential cross information can influence decisions. According to the authors, this kind of workflow prevents examiners from major problems such as the accessibility of contextual information that may to biased decisions. Regarding confirmation bias, the authors also suggested that forensic examiners should work within a blind setting, in order to avoid information which might not be relevant for their work.

The need for a verification phase was also mentioned. The verification phase should be carried once or twice after the primary analysis, and even using a cross-laboratory approach. Regarding this step, some disciplines such as fingerprint examination already integrate within their process a verification phase (see section 1.7.1. and 1.7.2.). The use of technology that complements decision-making (e.g. fingerprint, DNA or facial databases) is an important point to consider within laboratory procedures. However, precautions were raised due to the possible expectations and sources of confirmation biases that those tools can create within examiners' decisions. Finally, the authors claimed the need for a certified education, such as ongoing training in topics related to cognitive contaminations within forensic domains.

Regarding the fact that forensic sciences are at stake, one must consider the impact of an idea that suggests biases as something to avoid completely. As seen in previous sections, biases are possible to mitigate, but difficult to avoid totally (Tversky & Kahnemann, 1974). Furthermore, an investment in this field without the right corrections and adjustments might paralyse the field of forensics as Champod (2014) suggested the risks of having too much investment only in biases within forensic sciences. Firstly, the risk of having forensic scientists working within a "*blind and immune*" (p. 107) perspective from external sources of influence, such as the investigation teams. From this view, a paralysed forensic scientist cannot exchange information about a case with a colleague from another forensic discipline, which in the author's view, can produce positive outcomes complementing the work between forensic scientists and investigators. Secondly, the risk of treating the forensic scientist as a "*black box expert*" (p. 108). As mentioned, forensic sciences should move away from the black box paradigm that requires the forensic scientists to provide binary conclusions (i.e. yes or no) to the court through their reports. Instead, Champod proposes that the field of forensic sciences should start to be focussing more on systematic measurements of the traces (i.e. evidence).

As Kassin et. al. (2013) stated, there is a psychological effect that can disturb forensic examinations, however, the authors also observe the need for more research in this field in areas such as contextual information or pressures from outside the forensic laboratory. The authors also report some suggestions for the

laboratories themselves and the courts. For the forensic laboratories, the suggestions are based on the need to have training regarding the psychological effects on workflow such as perception, decision-making processes and social influence. Regarding the courts, Kassin et al. (2013) claimed that three possible problems might occur due to consequences of bias within forensic sciences. The first one states that biases can corrupt the conclusions and testimony of forensic examiners. Secondly, these corrupted conclusions will seriously affect future decisions since the court will continue to agree with the binary reports previously mentioned by Champod (2014), instead of challenging the evidence. Thirdly, Kassin et al. (2013) observed that legal professionals such as judges and lawyers may perceive forensic evidence in an overly reliable way. Thus, the authors also mention the need for education about the psychological effects for the professionals of the legal system when they face forensic evidence.

There are some limitations (both practical and theoretical) to the implementation of the recommendations provided in this section. One example is related to the suggestions for blind analysis in forensic laboratories (Kassin et al., 2013; Dror, 2013). It is possible to see the positive effects this methodology would have. However, if the focus is shifted to the motivation that forensic examiners have if they work in a completely blind setting, perhaps some negative points may be raised as well, as seen in previous sections which suggested feedback and contextual information sometimes have positive influences on the decision-making process. Another possible question is associated with the suggestion of having a cross-laboratory analysis. There is little doubt that this suggestion brings positive points to avoid sequential bias within the same forensic laboratory, however numerous cuts to the human resources of forensic laboratories have been made due to governmental strategies and other issues (Peachey, 2015), which may be a barrier to cross-laboratory verifications.

This section raised some important issues regarding forensic sciences, including fingerprint examinations. First of all, it is frankly of importance to take into consideration that even though forensic sciences have certain specifications within the tasks carried out, decisions that are made by professionals within this field are not different from the decisions made by medical staff, pilots, or other professionals that need to make decisions with greater impact on others' lives and which can also

suffer from biases. The examiners that work within the different disciplines of this field are influenced by the same type of biases, namely contextual and confirmation biases, that other experts are as seen in previous sections such as medicine or aviation. The stakes of observing errors within forensic sciences are high, emphasising the need for attention and to avoid errors. This need might be controversial. Whereas on one hand, society must see forensic sciences as a valid way to inform the decisions of the courts, on the other hand, a strong investment in blinding forensic sciences from external influences that may promote biases can become an obstacle to innovation and improvement. As Dror (2012) mentioned, risks and benefits of blinding forensic scientists need to be balanced. This thesis explores a variable (motivation) that may suffer from a blind setting, even though it might be key to the fingerprint examiners' performance. In the following section, similar issues are discussed, focusing only on one of the forensic disciplines, fingerprint examinations.

1.7. DECISION MAKING AND COGNITIVE BIASES IN FINGERPRINT ANALYSIS

Fingerprint analysis has been used to identify individuals for more than 100 years within the legal system (Earwaker, Charlton & Bleay, 2015; Cole, 2002). However, Barnes (2011) stated that "*friction ridge skin impressions were used as proof of a person's identity in China perhaps as early as 300 B.C., in Japan as early as A.D. 702, and in the United States since 1902*" (pp. 1). The term 'fingerprints' are part of a bigger field of work, which can be referred to as *ridge skin prints* since it refers to the study of a specific type of skin, the friction ridge skin and furrows (Smith & Bond, 2015). One of the first books on fingerprints was authored by one of the prominent scientists in the 19th century Sir Francis Galton. In his book 'Finger Prints', Galton (1892) started to differentiate what today the forensic community refers to as minutiae[2], which are formed by the arrangements within friction ridges in each fingertip. Friction ridges can be observed not just in the finger distal [phalanges and thumbs] but also in other parts of the human body such as palms, toes and soles of the feet (Champod & Chamberlain, 2009). The ridges themselves

[2] At the time these details were called Galton details.

form different patterns which are jointly aggregated in ridge systems, forming key focal points referred to as cores and deltas (Champod & Chamberlain, 2009).

There are differences in the frequencies of patterns that are observed. Smith and Bond (2015) described that the patterns that are most expected to be observed are loops (65%), followed by whorls (35%) and finally arches (5%) are the least frequently observed.

These points may differ in their organisation regarding the orientation of the ridges and their details. The organisation of the ridges and its details is referred to as minutiae. To examine and classify fingerprints, practitioners started to organise the structure of minutiae in three different levels – level 1, level 2 and level 3 of minutiae details (NIST, 2012):

- **Level 1 details**: Are defined as the *ridge flow*, i.e. the global pattern the print presents, which can be one of three different types [arch, loop or whorl]. At the first level of minutiae, it is also possible to observe the type of overall curvature the print has [to the right or the left].

Figure 16. Ridge flows patterns (arch, loop, whorl) retrieved from Champod and Chamberlain (2009)

- **Level 2 details**: Defined as the 'ridge path'. As seen previously, ridges can organise in different ways, forming different patterns (Champod & Chamberlain, 2009). Due to these differences, the forensic community addressed a number of ridge paths types, amongst them, one can find for instance of ridge endings (when a ridge comes to an end), bifurcations (one ridge splits in two), dots (a single dot also known as island), and lakes (also called 'open field' – a feature of minutiae absence).

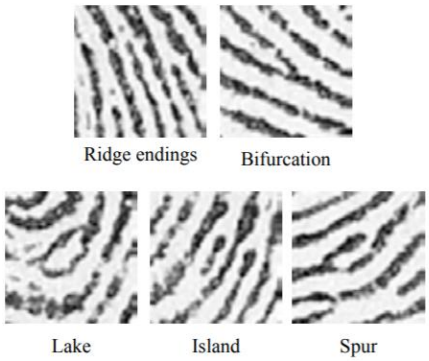

Figure 3. Examples of level 2 details retrieved from Champod and Chamberlain (2009)

- **Level 3 details**: Associated with the edges of the ridges and the pores. This third level of minutiae is not often extremely clear due to distortions (also known as 'noise') in crime scene marks, however, when present it can be a potential feature of identification to an individual.

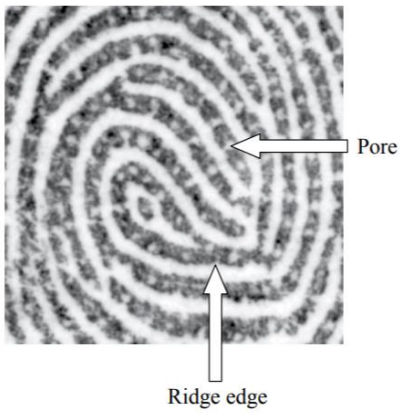

Figure 4. Examples of level 3 details retrieved from Champod and Chamberlain (2009)

Amongst the three levels of minutiae seen above, there is also the possibility to find fingerprints with some abnormal distortions. In these non-natural features, one can find scars of erosion on the fingertip itself or attempts to change its appearance. There are also some diseases such as some types of cancer that erode the ridges of the skin (Worland, 2015). There are also instances of individuals attempting to change their fingerprints in an effort to avoid identification. An example of this can be illustrated with the case of John Dillinger (Cummins, 1935) who successfully obliterated his fingerprints with acid. Although this was successful, the new features in his fingerprints were also seen as unique, making identification easier.

Figure 5. Oblitered fingerprints (scar and mutilation) retrieved from Yoon, Feng and Jain (2011)

To communicate the quality of ridge detail deposited on different surfaces the forensic community distinguishes the terminology in use. For a partial representation of characteristics left on a certain surface originating from friction ridges the scientific name used is *mark* (Champod & Chamberlain, 2009). When a mark requires a specific set of techniques to be visualised, collected and analysed, the forensic community uses the term 'latent mark' (Champod & Chamberlain, 2009). The quality of latent marks is also variable due to different features such as the surface where the mark was left [plastic, paper, metal, etc.] and the pressure made when deposited. Thus, "*quality is an assessment of accuracy of the representation of the impression*" (Champod & Chamberlain, 2009, pp. 59). Whereas prints tend to have a higher quality when they are taken using high quality techniques (e.g. rolling print paper or digital scan), crime scene marks may be of lower quality due to the factors described above.

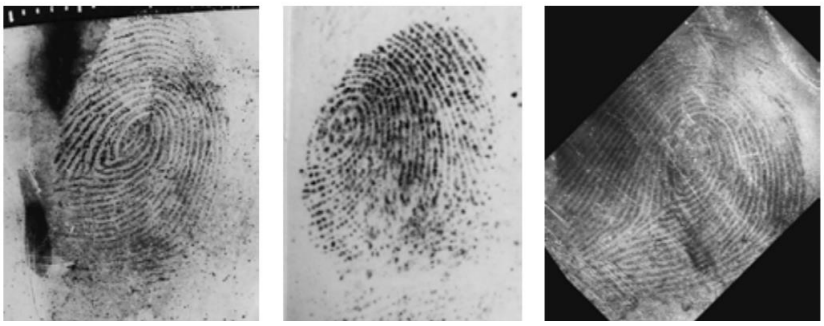

Figure 6. Fingermarks deposited in different surfaces (sheet of plastic, sheet of paper and adhesive side of a tape) retrieved from Champod and Chamberlain (2009)

With so many different features in the study of fingerprints, a system was needed to organise and search all the fingerprints in a system. Following Francis's Galton's work, Edward Henry organised the different types of minutiae, also known at the time as the 'Galton-Henry classification' (Champod & Chamberlain, 2009). This

system allowed the storage and manual retrieval of a particular set of fingerprints from bigger databases during an investigation. It was the beginning of what today is called the Automated Fingerprint Identification System(s) (AFIS). An AFIS system is a software that eliminates the need for a paper database with all the individuals digitally organised and searchable (Komarinski, 2005). An individual's fingerprints are collected (e.g. due to being suspect of a crime scene or regarding police intelligence) and then compared with the existing fingerprints in the database. Although there were some stand-alone AFIS systems in the past, Komarinski (2005) noted that these systems were now more likely to connect with related systems to contribute to a greater network of fingerprint identification systems. Since fingerprints are examined by human experts, one can assume that potential issues regarding human factors as seen throughout this thesis may arise. As it will be demonstrated, fingerprint examiners do suffer from biases. However, other influences can be observed. There are differences in guidelines between laboratories in different countries and therefore differences in the standards in place that may affect the quality of examiners' work (Gonçalves, 2017).

1.7.1. The ACE-V: Problems of terminology and reporting with quality

Within fingerprint examinations for one identification to occur a scientific process referred to as ACE-V (Analysis, Comparison, Evaluation and Verification) is required to be carried out by fingerprint examiners (ENFSI, 2015). An important issue to acknowledge is the fact that it seems that there is no global terminology with regards to the conclusions reported by fingerprint experts. Terms such as *identification*, *individualization* and *uniqueness* are often used, and somehow might be confused to mean the same thing, i.e., when an examiner concludes that a crime scene fingermark *matches*[3] a fingerprint of an individual. Even the European Network of Forensic Science Institutes (ENFSI) does not seem to have adopted agreed terminology as seen on its field-specific manual released in 2015 – Best Practice Manual for Fingerprint Examination (ENFSI, 2015). Within this manual, readers can find the definition of *identification* and *individualisation* to be very close (p. 56). An identification is claimed when a specific print passed through a

[3] Official terms are explained afterwards. *Matching* means, in a non-theoretical, professional and official way that both items (crime scene mark and known fingerprint) belong to the same person.

fingerprint comparison process and was attributed to an individual, whereas individualisation is claimed when a mark is attributed to a particular individual. For one individualisation, there needs to be sufficient quality and quantity of ridge flow, ridge characteristics and details in agreement with no unexplainable differences in the opinion of the practitioner.

Cole (2014) has criticized the use of terms such as "*individualisation*". The author claimed that such arguments were merely a "*desired reporting conclusion*" (p. 144) to the courts, instead of using a scientific framework terminology that "*flows logically from a justification*" (p. 144) such as identification. Saks and Koehler (2008) mentioned that forensic scientists should follow DNA analysis protocols to assess and report evidence. The main message these two authors shared was that forensic evidence should use a better and more elegant probabilistic methodology, even though they claimed that in some areas of forensics such as toolmarks, firearms or shoeprints, using probabilities could be more difficult. This method to report evidence was also shared by Champod (2009), who suggested likelihood ratios in order to make it clearer and more precise. Champod criticised the fact that forensic scientists after the release of the NAS Report (NAS, 2009) were, in a fair majority, used to report evidence without using a functional set of logic. Besides that, the use of likelihood ratios, acknowledges the strength of evidence, without dismissing the fact that there could be more evidence not known at the time of analysis. The use of likelihood ratios would promote the adoption of a convention "*that allows the scientific statement to remain consistent within a given framework irrespective of other evidence at hand*" (Champod, 2009, p. 3).

Kaye (2009) attempted to clarify the difference between individualization and uniqueness. According to the author, forensic examiners can claim individualization without claiming uniqueness. Individualization was acknowledged by Kaye as the quality characteristics that are needed be observed to claim an item was originated from the same source, in opposition to other class characteristics that suggest an item (or another source of intelligence) to be unique. This challenged some of the NAS report (NAS, 2009) contents, specifically when it mentioned that any forensic testimony or reporting should include clear characterisations of the limitations present as well as measures of those limitations. However, Kaye (2009) argued that there was no evidence that numerical

presentations were more comprehensible than qualitative ones (p. 1177). Therefore, there was no concise and agreeable way to report forensic evidence that would completely fulfil court requirements.

One of the landmarks within fingerprint identifications was the release of the NAS Report (NAS, 2009). In this report, claims regarding the subjectivity and validity of the methodologies that fingerprint examiners were using, were challenged. In order to regulate as well as to calm and support the forensic community specifically, and society in general, large professional bodies responded to the report.

Regarding terminology in the field of forensics, particularly in fingerprints, there is a recently published online lexicon by the Organization of Scientific Areas Committees (OSAC), aiming to make forensic examiners aware of the importance of using the same language across the discipline and casework. This manual was created as different fields use the same terminology but with different meanings, as stated by the National Institute of Standards and Technology: "*To help facilitate clear communication across the many disciplines*" (NIST, 2018).

The Forensic Science Special Interest Group also have mentioned the importance of implementing quality standards (Earwaker, Charlton & Bleay, 2015). In this document, authors mentioned that within the great challenge of being accredited, fingerprint bureaux would need to consider standards that were referred in this document, and that were related to different aspects of fingerprint practice such as technology that should be robust enough to "*stand up to tough scrutiny in a court of law as well as satisfy the scrutiny of validation under future ISO 17025 accreditation which is expected to be introduced for fingerprint bureaus in the UK in 2018*" (p. 11). Throughout the document, authors also presented some of the professional bodies that have been promoting best practices within the fingerprint community such as the International Association of Identification (IAI) or the European Network of Forensic Science Institutes (ENFSI).

The IAI, which is the largest professional body of fingerprint professionals, and the old Scientific Working Group on Friction Ridge Analysis Study and Technology (SWGFAST), currently Organization of Scientific Areas Committees Friction Ridges Subcommittee (OSAC FSR), which is the body that aims to establish and

suggest guidelines for the practice of friction ridge examiners, have been working to common objectives. The IAI has been delivering certification in different types of disciplines where one finds fingerprints. The IAI and the OSAC bodies also suggested fingerprint professionals to try avoiding the use of terms that could be less accurate such as 'zero error rate' as mentioned by Kaye (2009).

The ENFSI also commented on the NAS report through the European Fingerprint Working Group (ENFSI EPWG) in a paper by Meuwly (2011). In this paper, the author argued that the position of the European group of ENFSI suggested that fingerprint evidence should not be assumed as 100% 'absolute' and capable to exclude individuals with total certainty. The OSAC Friction Ridges subcommittee attempted to formalise the term individualization, which is the terminology suggested for use by official guidelines. They suggested that an individualization should be perceived as "*the decision by an examiner that there are sufficient features in agreement to conclude that two areas of friction ridge impressions originated from the same source. Individualization of an impression to one source is the decision that the likelihood the impression was made by another (different) source is so remote that it is considered as a practical impossibility*" (SWGFAST, 2011, p. 1). Here, one can observe the inclusion that Champod (2009) suggested regarding the use of likelihood ratios.

The discussion related to the terminology that has been used within the fingerprint community is of great interest and importance. All of the terms (e.g. individualization, unique, identification) that have been discussed throughout this section have been discussed by organisations that have been working towards the improvement of fingerprint examiners' practices. In the next section, this thesis will focus on the main process that fingerprint examiners follow during their work, namely the ACE-V, and how its design influences fingerprint practice.

1.7.2. The ACE-V: method or process?

According to the scientific guidelines (FSR, 2011; NIST, 2012; SWGFAST, 2002) for the discipline of fingerprints, the ACE-V process is composed of four different phases, each requiring decision making with different influencing variables.

Figure 7. The four phases of ACE-V process

The analysis phase is an initial information-gathering where fingerprint examiners study the information present in the unknown/crime scene print in terms of quality and quantity of details. Throughout this phase, examiners need to take into consideration information such as the substrate, development method, and the levels of ridge details present as well as distortions.

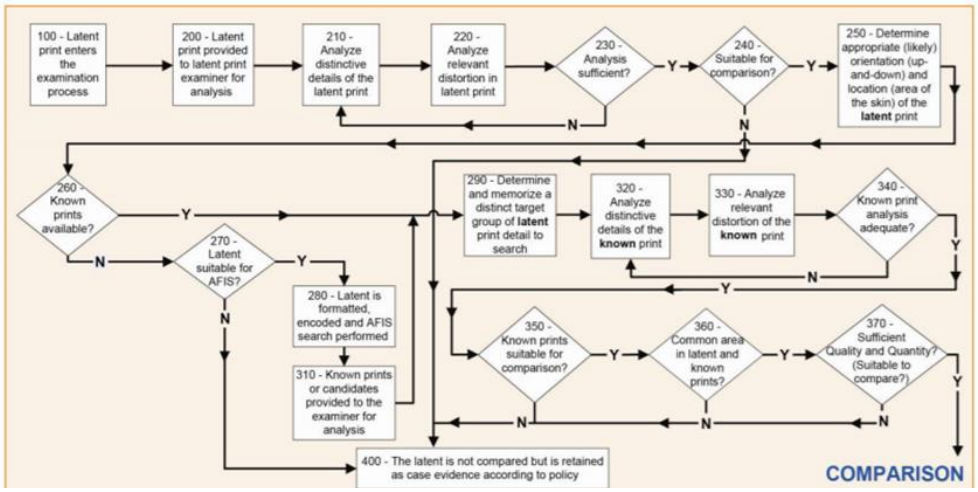

Figure 8. Analysis phase of ACE-V

After this phase, examiners shall have their comparison regarding the usability of the evidence in terms of its quality (if the evidence is good enough to be used) and the quantity of details (i.e. the number of minutiae that is possible to observe) in order to carry out the further phases of the process. Comparison happens when fingerprint examiners compare the unknown fingerprint with possible candidates provided by the AFIS system. This comparison is usually made in a side-by-side way. In this phase, fingerprint examiners compare the quantity of agreement and disagreement of ridge details in both fingerprints (unknown and candidates given by AFIS) and determine which of the candidates have sufficient characteristics to carry out the next phase.

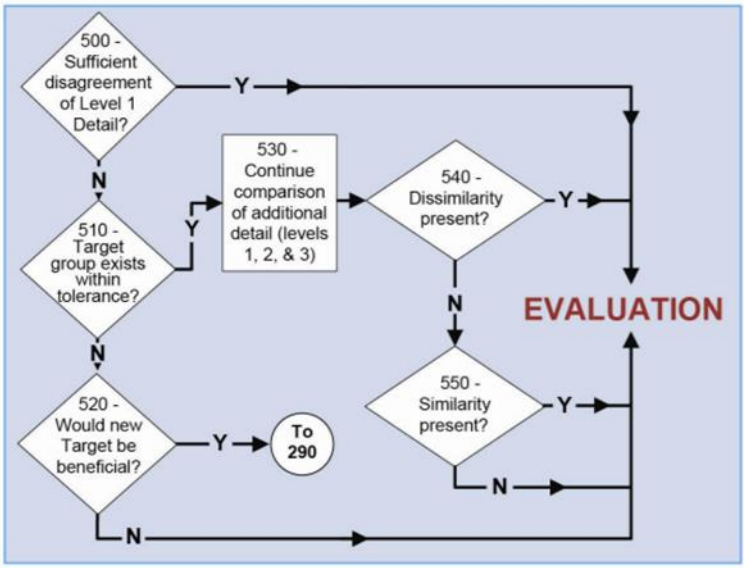

Figure 9. Comparison phase of ACE-V

The Evaluation phase is the last phase of the ACE set. It is supposed to be "*the formulation of a conclusion based upon analysis and comparison of friction ridge skin*" (SWGFAST, 2002, p.3). During the evaluation, fingerprint examiners assess the number of ridge details in the unknown fingerprint and the chosen fingerprint candidate given by an AFIS system (or similar) in the previous phase. There, they must reach a conclusion. Conclusions for a match (i.e. both prints were originated from the same source) need to be drawn from more than first level details in agreement. Thus, matches are concluded when examiners can identify sufficient agreement of first and second level details (or even third level details). Besides an individualization (i.e. a match), examiners can also conclude other options such as an exclusion or an inconclusive decision. Exclusions are made when examiners are able to reach enough first, second and sometimes third level details in disagreement. If examiners are unable to determine a number of details in disagreement, an inconclusive decision shall be warranted (SWGFAST, 2002).

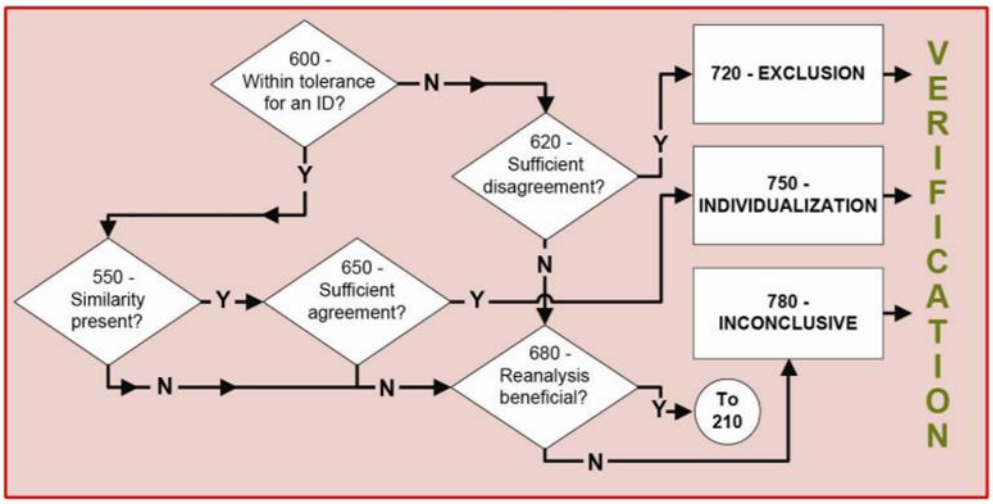

Figure 10. Evaluation phase of ACE-V

After the ACE set is done, the Verification phase can start. Professional guidelines such as the notes that SWGFAST has published state that this phase involves a review of the work made by the fingerprint examiner in the previous phases conducted by a second fingerprint examiner. In some laboratories, the conclusion reached (match, non-match or inconclusive) is known to the verifier. However, other laboratories keep this information confidential for a second examiner to review the previous work independently. UK guidelines require one verification for individualisations (matches) and rejections (non-matches) by a second examiner (FSR, 2011). Even though the process might seem simple as laypeople see the entire process encompassed within four phases only, the guidelines (NIST, 2012) described the entire ACE-V with more than 50 steps between its start and reaching a conclusion.

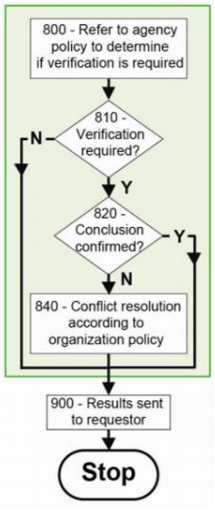

Figure 11. Verification phase of ACE-V

It is important to note two issues regarding the ACE-V methodology. The first one is the complexity of the process. Even though figure 7 (p. 82) is simplified, the process itself can be very complex and extremely time-consuming. The second issue relates to the phase which will be focused on in more depth in this thesis, the verification phase. As seen, the verification phase can be made independently or with knowledge of the previous conclusion reached by the initial fingerprint examiner. This is of interest for the empirical chapters in this thesis, in order to understand if there are problems such as contextual and confirmation bias within verification procedures which are not independent (Kassin et al., 2013; Tversky & Kahneman, 1974). It is also important to understand what fingerprint examiners think about suggestions regarding independent procedures such as blind verifications (Dror, 2013).

Regarding the verification phase, the SWGFAST (2011) suggested this phase to promote quality assurance (QA) protocols and procedures for the friction ridge examinations. According to this working group, the verification phase should be implemented within a blind setting. The first objective with this suggestion was the attempt to promote reproducibility of the conclusions that fingerprint examiners achieve during their case-work. According to the guidelines, the blind verification should be an independent phase, as mentioned above, where an independent examiner, i.e. an examiner that did not join the team that executed the first three phases (ACE), has no expectation regarding the possible conclusion of the case. The reason why the guidelines stated that the blind examiner (or the verifier) has no expectation is due to the fact that the guidelines suggested that these blind verifiers should work without any contextual information or other information regarding the previous examiners' conclusions. The working group also stated that blind verifiers can be used in all of the phases of the ACE method as well as in a variety of specific situations, such as (1) comparisons where there is a high level of distortion in the fingermark, (2) comparisons with a low number of minutiae, or where minutiae has low quality, (3) when there is the possibility of analysing two fingermarks at the same time on the same surface, (4) to clear conflicts among examiners, (5) when there is a large disparity between the experience of the first examiner and the verifier, (6) when fingermarks are retrieved from probative locations such as weapons, (7) cases where the evidence is based on a single fingermark, (8) if the origin of the ridge details is not certain, i.e. it may have

originated from a palm, hand or toe or any "*other circumstances at the discretion of the examiner*" (p. 2). Their recommendations suggested that 10% or 25 cases per examiner should be blind verified, or "at agency discretion" (p. 2). The guidelines also defined the characteristics of the examiners who conduct blind verifications. They need to be trained to competency and they should not have previously consulted with examiners that were involved in the case.

The ACE-V methodology has been the focus of research articles, which have considered the merits of this examination method. Triplett (2012) considered whether the ACE-V approach is a process or a method. Triplett began to analyse this by differentiating the concepts of 'method' and 'process'. In her perspective, a process can be considered as a *task* which needs to be performed. The method is, therefore, the procedure ("*specific set of actions*") one does aiming to achieve the goal of completing the task (Triplett, 2012, pp. 6). Regarding the procedures of ACE-V, it was suggested that each phase can be observed as a process, therefore concluding that ACE-V is a set of processes. However, each laboratory seems to employ its own specific methodology based on the standards their agency adopted. The use of different methodologies in the same process (e.g. using magnifiers or digital software), might be one reason why, for the same fingerprint comparison, some analysts conclude to an individualization, while others may reach inconclusive decisions (Triplett, 2012).

It is, therefore, useful to consider some of the differences between national agencies in different countries regarding the stated guidelines for fingerprint analysis. In the United Kingdom, the Forensic Science Regulator published a set of official guidelines (FSR, 2017, 2011) which include suggestions for quality standards within fingerprint analysis using ACE-V. It includes acknowledgements of the complexity in fingerprint analysis processes as well as the experience and training that an examiner goes through to achieve this kind of expertise. The process of comparing ridge details also received attention in these official documents. In the past, there was a minimum of sixteen ridge details that should be in agreement between the unknown fingerprint and the candidate selected from the AFIS system (Evett & Williams, 1996), however, since 2001 the UK moved from a numerical to a non-numerical standard where current guidelines suggest the comparison be made without this minimum number of sixteen ridge details in agreement, instead, the

process relies in the quality and quantity of each comparison which is then verified (FSR, 2013, 2011). This change was made right after the analysis of a case that observed media coverage due to its characteristics (the Shirley McKie's case – described in the following section 1.7.3.). After the closing of this specific case, the Scottish Executive opened an inquiry led by Sir Anthony Campbell (The Fingerprint Inquiry, 2011) which took place between 2009 and 2011, and which eventually had an impact on the standard used within UK forensic bureaus as from its publication it was observed the abolishment of the 16 point standard.

Similarly, in the United States of America, the National Institute of Standards and Technology [NIST] published the report Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach (NIST, 2012) suggesting that some laboratories may require their examiners to document all features during the analysis phase if the unknown fingerprint has less than 12 ridge details. In a slightly different way in Germany, the national agency Bundeskriminalamt published its guidelines for standards in fingerprint examinations. One of the main differences from the two previous cited reports is the fact that German examiners need to reach at least 12 minutiae points of agreement during the comparison, or if 12 points are not in agreement between the latent print and the known print, quality standards such as the third level of minutiae need to be included in the analysis (Bundeskriminalamt, 2010). The federal agency in Portugal (Portuguese Judiciary Police) suggests Portuguese fingerprint examiners should identify 12 points of minutiae plus an extra point (called security point of minutiae) to conclude an identification (Correia & Pinheiro, 2013). Other countries also have different thresholds, the previous examples were only to illustrate what contributors from the FBI and the SWGFAST on the NAS report (2009) mentioned. According to the NAS report, the threshold that fingerprint examiners use to express an identification, or any other sort of conclusion is rather subjective. Thus, it is important to consider the differences between laboratories that work under different standards (Gonçalves, 2017) and acknowledge the fact that this can be indeed a variable that besides making collaboration harder, may suggest that the fingerprint examination process as a whole is rather subjective and therefore able to be questioned in its method.

Neumann (2012) described the two approaches used within fingerprint bureaus, the holistic approach and the numerical approach. Although the numerical approach was the first type of standard to be implemented, that standard developed differently in countries such as the U.S. Aiming to be more approximated to Galton's philosophy (1982), this approach was used informally by examiners until the IAI in 1973 commented that using a numerical standard to assess fingerprints and to make comparisons was scientifically questionable (Champod & Chamberlain, 2013). Since then, some fingerprint bureaus adopted the holistic approach which suggests fingerprint examiners should compare the friction ridge details within a fingermark and a candidate print in a continuous manner using the quality of each piece of comparison rather than counting points of agreement (Neumann, 2012).

The numerical approach requires a certain number of minutiae points to match in order to claim that a fingermark is a match with a candidate fingerprint. Using this numerical standard there were three possibilities to classify a comparison. If a comparison found more than 12 minutiae points that matched, claiming a match was "*beyond debate*" (p. 228). If the comparison found between 8 and 12 minutiae points, there could not be claimed a match with certainty, and additional assessments should be made such as "*(a) the quality of the fingerprint, (b) the rarity of the minutiae, (c) the presence of a core and delta in a clear area of the print, (d) the presence of pores, and (e) the perfect agreement of the width of the ridges and furrows, the direction of the ridge flows, and the angular value of the bifurcation*" (p. 228). Finally, if the comparison found "*a limited number of characteristics*" (p. 228), the comparison cannot be claimed as a match with any certainty. As discussed previously, even though the numerical standard was well implemented in the past, it can differ between countries (Neumann, 2012). However, there were countries like the United Kingdom that had their numerical standard challenged. Evett and Williams (1996) argued that there was no statistical reason to keep using a 16-point minutiae standard to claim a comparison to be a match with certainty, and therefore, a holistic approach, described above, was adopted for the ACE-V process in bureaus within these countries.

It seems that each phase of ACE-V is a process (i.e. a set of actions). However, when examiners perform it according to specific and validated guidelines, ACE-V is a method. Nevertheless, different methodologies can be used for the same

process, and previous research has not demonstrated which approach is most effective. Regarding the differences between countries, it is necessary to focus on how examiners operate with the different guidelines they receive and to what extent they perceive and agree with these guidelines. Another central issue is the importance of the verification phase and also the setting where one performs it in a blind way (blind verification). As mentioned by a number of official guidelines (NAS, NIST, FSR, SWGFAST) and authors (Kassin, Dror & Kukucka, 2013; Champod, Lennard, Margot & Stoilovic, 2016), the verification phase should be implemented within conditions that allow examiners to carry out the ACE process phases independently and without any kind of disturbance. An ideal setting for that would be to implement blind verification for all cases. However, financial costs and organisational structure can be an obstacle to implementing that successfully (Gibb, 2019).

As far as one understands better the ACE-V process, it is possible to observe the links with the work cited in section 1.2. "Decision-Making Processes" where Kerstholt and Raaijmakers (1997) defined dynamic tasks. In each phase of the ACE-V examiners are confronted with steps that can be dynamic, i.e. that are dependent from a number of variables such as the urgency of the examination (for instance if the examination is from a serious crime) or the availability of resources (e.g. are there enough human resources to conduct a verification?). When conducting the several steps that are integrated within the ACE-V process, examiners need to have mental representations of the process as a whole in order to understand not only its sequence but to ensure it follows a chain of custody that prevents errors throughout the process. Another relevant variable that Kerstholt and Raaijmakers (1997) referred was feedback which as seen in a previous section dedicated to this concept (section 1.5.2. Feedback in Decision Making) it plays a great role within any type of decision process where fingerprint examinations are not an exception.

1.7.3. Cognitive Biases within fingerprint examination

Fingerprint examination has been one of the most widely utilised evidence types presented in court for more than a century (Cole, 2002), and confidence in this method has not been questioned much in the past by the public and judicial system

(Charlton, Fraser-Mackenzie and Dror, 2010). OSAC stated some of the forensic disciplines where more research was needed, and in 2015, fingerprint analysis and the ACE-V procedure was the focus of an assessment for research (NIST, n.d.). One of the most influential cases that highlighted errors based on confirmatory bias within fingerprint examinations, which received wide media coverage and attracted the attention of the criminal justice system and researchers, is the case of Brandon Mayfield – the Madrid Bombing Case (Thompson & Cole, 2005; Kershaw & Lichtblau, 2004). In this case, four FBI latent print examiners incorrectly identified Brandon Mayfield as the person to which the fingermarks found on items related to the terrorist attack belonged. The analysis of the fingerprints (unknown and candidates from AFIS) was made independently according to the report provided by the Office of the Inspector General (OIG, 2006). The report stated that latent print examiners were not aware of variables that could have affected their performance such as the religion of Brandon Mayfield. The report concluded that the main cause of the misguided identification happened due to "*circular reasoning*" (pp. 7) by the examiners, i.e. the analysis started with the normal ACE-V procedure, but since only 10 ridge details were found, examiners did part of the analysis using backward reasoning trying to find ridge details from Mayfield's fingerprint (given by AFIS) in the unknown fingerprint. After recognising the wrongful decisions made, the FBI publicly apologised to Mayfield in a statement (FBI National Press Office, 2004) mentioning that the fingermark used during the investigation was based on an image of substandard quality.

The review made by the OIG (2006) on the Mayfield case claimed that there was no evidence of intentional misconduct either by laboratory examiners or by inspectors leading the criminal investigation. The American bureau also stated that the FBI's Latent Fingerprint Unit was going to review their practices to consider the adoption of new guidelines in their procedures. Due to the intense media coverage that this case received, research in this area started to have a bigger impact and promoted the development of documents such as the NAS report (2009) already discussed in previous sections.

In the UK, another case of confirmation bias attracted media attention; the Shirley McKie case. McKie was accused of perjury in 1997 after a fingerprint found at a crime scene was wrongfully identified as hers (BBC, 2011). At the time of arrest,

Shirley McKie was a police constable who testified that she did not enter the crime scene in this case, but the fingerprint examiners working at the time claimed to have identified her fingerprint on a door frame in the crime scene. McKie was charged with perjury based on testimony she had given and in May 1999 she was cleared of her charges at the high court in Glasgow (O'Neill, 2011). To be cleared of the charges, McKie had fingerprint examiners from different agencies around the world to challenge the validity of the evidence (Charlton, Fraser-Mackenzie & Dror, 2010). After McKie's charges and eventual exoneration, the Scottish Executive opened an inquiry led by Sir Anthony Campbell (The Fingerprint Inquiry, 2011) which took place between 2009 and 2011.

The inquiry presented 10 key findings and a total of 86 key recommendations. Amongst the findings and recommendations, there were important messages that Sir Anthony Campbell stated such as "*fingerprint evidence should be recognised as opinion evidence*" and that evidence could not be accepted with "*100% certainty*" (The Fingerprint Inquiry, 2011, pp. 740). Besides the abolishment of the 16 point standard (mentioned in the previous section 1.7.2.), the inquiry also found that fingerprint examiners should keep engaged either in on-going training as well as with academic communities to improve their practices, which is comparable to what the American reports (e.g. NAS Report, NIST Report) concluded.

There is an interaction between human and technology when latent print examiners make use of tools such as AFIS. This link between humans and technology can also be the source of errors based on confirmation biases. Dror, Wertheim, Fraser-Mackenzie and Walajtys (2012) designed an experiment in which 23 latent print examiners (11 were IAI certified) received their workflow as if it was a regular working day, and as usual they would use the AFIS system to generate possible candidates to compare with the latent print from the crime scene. As previously described, AFIS provides examiners with a set of potential candidates within a ranking where the more likely a fingermark is to be the match with the latent mark from the crime scene, the higher it appears on the ranked list provided. However, in this study, the authors manipulated the ranking that AFIS provided to the participants and therefore the candidate prints were given in a random order of similarity to the examiners. A total of 55,200 comparisons was carried out by participants with 1832 match comparisons being made. From these, a total of 1516

errors of all types were observed. Errors were of three types, false identification (n = 49), false inconclusive (n = 1001) and missed identification (n = 502). Another factor the authors observed was the response time that participants took to make the comparisons. The authors noted that examiners who spent less time on a potential candidate that was in a higher position due to the misleading AFIS ranking were more prone to commit an error ($p < .001$). These results showed that participants wrongly chose the potential candidates which were presented in higher positions of the deceptive AFIS ranking and that they spent less time on these. It appears that although technology could be a way to facilitate some parts of the examiners' work, it might have within its use some risks. Examiners were not aware of this source of bias which may contribute to false identifications when candidates were positioned more closely to the top of the list of AFIS, showing the effect of a confirmation bias within the position of each candidate in the set given by the system.

A study that was central for the empirical work described in Chapters 3 and 4 of this thesis was the work carried out by Langenburg, Champod and Wertheim (2009), who observed the effects of contextual information within the field of fingerprints, and how this kind of information could influence fingerprint examiners' performance. The study carried out at an International Association of Identification Conference had four main objectives: (1) understand the influence that the reputation of the examiner who carried the ACE phases had on the participants' performance, (2) explore variables such as training, education and experience, (3) conduct tests with non-professionals to challenge their results with a control group as a baseline and finally (4) understand if there is a need to suggest blind testing exercises. To do so, the authors invited 43 fingerprint examiners (experts) and 86 novices (individuals that did not have any experience analysing fingerprints) to complete the experiment. To test contextual influence, authors deceived participants by mentioning that the experiment was to measure the variation of examiners' evaluation during latent print comparisons. Three groups were created in both pools of participants and then randomly assigned to perform 6 comparisons either on paper or digitally. Comparisons could have three types of conclusions; individualization (same source), exclusion (different sources) or to be inconclusive. Definitions of these conclusions were provided to standardise the understanding of the terminology.

One group had a high rate of contextual information where it was mentioned that an internationally recognised fingerprint examiner conducted a prior comparison, a second group had a low rate of contextual information, where the information is given anonymously, and the third group was used as the control group as participants in that group did not receive any contextual information related to the previous work before the verification phase (NIST, 2012).

Laypeople's results showed a contextual bias effect. In the control condition, only 6% of the novices' responses followed the prior conclusion that was supposedly conducted by the previous examiner, in the low bias condition this effect increased between 1-7% (authors considered it non-significant), however, in the high context condition results yield a significant 32% of respondents to be affected by the bias prompt (i.e. they followed the prior conclusion). In the experts group, an effect was also reported by authors, however, there was a difference to the novices' responses. While the novices increased their wrong decisions due to misleading contextual information, fingerprint examiners showed a more conservative approach by giving more inconclusive answers when their opinion was different from the contextual information they were provided with. This study demonstrated the risk of fingerprint examiners having information during the ACE-V process and to work within non-sequential (i.e. linear) conditions as suggested by Kassin, Dror and Kukucka (2013).

Still related to the effects that contextual information has within examiners' performance, a study conducted by Earwaker, Morgan, Harris and Hall (2015) demonstrated that different types of context will affect how fingerprint examiners carry out submissions of evidence. Their study asked 11 participants to decide about submission decisions in relation to borderline quality fingermarks. Participants were split into two groups; one group received first stimuli related to "serious crime" and afterwards stimuli related to volume crime and another group that received stimuli in the opposite way. Results showed there was a significant relationship between crime context and the outcome of submitting a piece of evidence ((x^2, 2) = 9.817, $p < .01$). Their findings found that fingerprint examiners treated fingermarks associated with different contextual information differently, even though guidelines suggest examiners to equally treat a piece of evidence from

a serious or a volume crime. This claim suggests that there are emotional influences regarding the process of fingermark recovery.

Tangen, Thompson and McCarthy (2011) identified what they referred to as 'the fingerprint expertise'. To achieve that claim, authors presented to two groups of 37 participants (experts and novices) pairs of prints displayed side by side on a computer screen and asked participants to judge whether the pairs matched or not within a confidence scale of 1 (certain of a non-match) to 12 (certain of a match). The task was made as a two-alternative forced-choice design, i.e., participants could not leave their answer as inconclusive. This allowed authors to distinguish between accuracy and response bias. Stimuli consisted of 36 simulated crime-scene prints paired with fully rolled prints. From the 36 pairs to judge, 12 prints were paired with a matching print (same source), 12 with a nonmatching print but similar to the given one (similar distractor) and 12 with a random nonmatching (non-similar distractor). Pairs were allocated in a random order to each participant. The results showed that novices were less accurate than experts in all three conditions. In the condition using matching pairs, experts judged over 90% of the pairs correctly (92.12%) and novices only achieved 74.55% correct answers. In the nonmatching condition with similar distractors, novices got less than half the pairs correct (44.82%) whereas experts scored almost all correctly (99.32%). Finally, in the nonmatching condition with non-similar distractors experts scored all the pairs correctly (100%) and novices were correct 77.03% of the time. Results were then subjected to a 2 (type of expertise) x 3 (type of condition) mixed analysis, which showed that there was an effect of expertise that explained accuracy in all conditions ($p < .001$).

The authors concluded that fingerprint examiners did have sufficient expertise to be distinguished from lay-people. However, more than acknowledging that fingerprint examiners commit errors throughout their practice, it was claimed that an important further step should be how to acknowledge errors. The same authors carried out another study (Thompson, Tangen & McCarthy, 2014) attempting to address this issue by observing the performance of latent print examiners with different levels of experience. Their sample was composed of 54 fingerprint examiners (37 qualified experienced examiners, 8 intermediate trainees, 9 new trainees) and 37 laypeople. In the experimental setting, examiners were presented

with pairs of prints displayed side-by-side on a computer screen and asked if the pairs matched or not in order to assess how examiners with different experience levels would evaluate the quality of information within each pair of prints. To answer this question participants needed to reply with a confidence rating between 1 (the marks are definitely different) and 12 (the marks are definitely the same). Aiming to distinguish between accuracy and response bias, a similar design to the previous study was used requiring participants to respond in a two-forced choice approach. The stimuli were composed of 45 prints taken from casework set for training purposes within forensic laboratories in Australia and were divided into three different categories (15 matching prints, 15 similar non-matching prints and 15 non-matching prints). After this clarification, the authors presented 537 matching trials, 547 similar nonmatching trials and 544 non-similar nonmatching trials. Pairs were all verified by three senior examiners and contextual information related to each pair were not available to the participants. There were four possible outcomes of each trial which are illustrated in the table below:

	Prints Match	Prints do not match
Examiner says match	Hit	False Alarm
Examiner says non-match	Miss	Correct Rejection

Table 6 – Possible outcomes within Thompson, Tangen & McCarthy (2014)

Results showed that experienced examiners and intermediate trainees had higher accuracy than new trainees and laypeople in all conditions. Experienced examiners and intermediate trainees responded closer to the extreme ends of the confidence scale compared to the two other types of examiners in the sample. In addition, experienced examiners did not significantly differ from novice examiners regarding the skill of matching fingerprints which are from the same source, however, experienced examiners had a higher accuracy when the fingerprints were similar and non-similar from different sources. A mixed ANOVA revealed a significant effect of expertise on accuracy ($p < .001$). To analyse accuracy levels, authors used a signal detection method and found that experienced examiners and intermediate trainees tended to be more conservative regardless of whether the pairs of prints matched or not. This suggested that examiners who have more experience also have a tendency to overcome the possibility of committing an error of having a false alarm due to a conservative response bias, however, the authors also claimed that

this behaviour promotes the possibility to observe a larger number of misses (when prints match and examiners say they do not). The authors discussed the potential trade-off of the cost that experienced examiners and intermediate trainees observe when pointing out a match that is a false alarm or even a miss compared with trainees and novices. In this study, besides the distinction that authors were able to make between experienced examiners and non-experienced examiners, the power and impact of response bias were clear. When responses suffer from response bias, conservatism will be lower, but also the possibility to commit errors that can be costly. This thesis will be referring to this trade-off in the discussion associated with another variable explored in the empirical chapters, motivation.

1.7.4. Fingerprint examiners' motivation

As demonstrated by previous research, the information received by examiners and the environment where these professionals work can influence their performance, potentially contributing to biases and errors. Although there are some published suggestions from researchers and official reports from national agencies (e.g. NIST, FSR) related to the ACE-V process and the cognitive contaminations within its decision-making processes, it seems that when it comes to what examiners think about these sorts of obstacles, there is a lack of research.

An exception to this is an article by Charlton, Fraser-Mackenzie and Dror (2010), which focused on the emotional experiences and motivating factors of fingerprint examiners. In this study, 13 fingerprint examiners (fully trained) with at least seven years of experience were interviewed. The contents of each semi-structured interview were divided into three topics: day-to-day fingerprint analysis processes, particularly harrowing or difficult cases and the role of being a fingerprint examiner. The data was analysed through a thematic analysis protocol and results identified five recurring themes: reward, motivation, satisfaction, fear and need for closure. These themes were then further divided into six sub-categories: job satisfaction and pride associated with using skills, motivation, satisfaction and hope associated with catching criminals and solving crimes, expression of satisfaction and motivation associated with the importance of the case, feelings towards searching and finding an identification [match] and finally expressions indicating a need for closure on case-work. The authors argued that there was no baseline to

compare their results with, which suggested that topics such as the ones explored in their study as well as the kind of methodology employed, should be further explored in order to verify reliability. In addition, the authors observed that during the explanation that participants gave regarding the identification process of fingerprints, objectivity was verified in respondents' discourse. However, that objectivity contrasted with emotive language and motivations observed.

It seems that the results of this study indicated the relevance of focusing on motivational factors within the fingerprint examiners' work. There are not many studies focusing on this topic which is something addressed by this thesis. Charlton, Fraser-Mackenzie and Dror (2010) study results found that the major motivational factors observed were related to identifying criminals and making correct identifications, which seem to be associated with the normative oriented motivations that Hoffman (2015) identified. These factors appeared as leading emotional factors regarding examiners' pride in their job. However, this feeling of being proud and motivated due to the context of confirming identifications can also be a source of bias, namely contextual and confirmation bias, particularly during high-profile and serious crimes, for instance, to solve a case as the Madrid Bombing Case.

As this section has demonstrated, the process that is necessary to carry out a fingerprint comparison is greater than just observing mere differences and/or similarities. Fingerprint examiners are required to complete a set of processes within the ACE-V methodology that is composed of 4 phases. When the analysis is initiated, the examiner needs to be accurate enough when identifying points of interest to then compare the crime scene fingermark to a set of potential candidates. Once the analysis and comparison are complete, the examiner will make his/her judgement, which should be based on material evidence, if possible using non-subjective forms of measure such as likelihood ratios. Technology has been assisting with these tasks. However, human performance is still the crucial tool to perform the assessment and to communicate the conclusions. Regarding the review of the task itself, the sections above elucidated the need to have a strong and coherent revision, referred to as the verification phase. One important aspect of the verification phase is the need to carry it out within a blind context, similarly to what should be done with proficiency testing. However, lack of resources and other types

of constraints (eg. time, technology, etc.) makes this difficult to implement. As professionals, fingerprint examiners have more knowledge and skills than laypeople. They perform better in tasks of identifying fingerprint matches and non-matches and they also have more rigour in terms of differentiating close matches and close non-matches. Nevertheless, they are subjected to errors, namely errors originated by cognitive biases. And due to that, official guidelines have been produced. The literature in this area still needs to focus on some aspects of the fingerprint identification process, specifically the verification phase as well as the emotional variables such as the motivation that fingerprint examiners must have to carry their job. The empirical work presented in this thesis extends the previous literature and makes an original contribution by exploring these variables further.

1.8. CONCLUSION

This chapter focused on the research literature that forms the basis for the topics explored in this thesis. As observed in sections 1.1 to 1.4., the literature was focused on general decision making theory and research. It was important in order to provide a general overview of the concepts, terminologies and research background for the specific object of this thesis, which was to explore the decision-making processes of fingerprint examiners. Therefore, sections 1.6. and 1.7. addressed the central topics but exclusively within the field of forensics, and then specifically fingerprint analysis.

Throughout the chapter, it was presented the features that can be observed within the process of making a decision such as the framing that initiates such tasks. It was observed that there are different types of decisions, some that are conscious and others that are automatic. One specific type of task was the focus of this introduction (dynamic tasks) as it was of particular interest in the type of decisions that fingerprint examiners conduct in their daily work. Dynamic tasks were described as sets of actions that are conducted within environments which can change due to variables such as the context, which is what happens in fields of forensic such as fingerprint analysis.

The literature review also described how experts differ from laypeople. This differentiation is of particular interest as fingerprint examiners have been the focus

of a number of studies that have challenged their level of expertise and their propensity to make flawed decisions. Expertise in this thesis was defined according to one specific line of thought, the Performance-Based Approach. Within its model, PBA defines individuals who are considered experts to be within a continuum of progress where judgement is a core skill they keep developing over their practice. Since fingerprint examiners' work is based in their judgement applied to a specific methodology, the Performance-Based Approach was chosen and used to explore results in empirical chapters where participants' accuracy are measured based in their judgement related to their experience (in the case of examiners) or the lack of it (in the case of laypeople) when conducting pattern recognition tasks.

Followed by the differentiation of expertise and non-expertise, a central topic of this literature review cognitive bias and its influences on decision making. Cognitive bias has been the focus of a large amount of literature that studied decision-making processes in general and also within forensics. In section 1.4. two specific types of cognitive bias were considered, the confirmation and contextual bias. Both types of bias have been considered to be rooted in decisions affected by heuristics, i.e., decisions that suffer from unconscious and automatic processes that often are flawed and promote erroneous outcomes.

One of the assumptions that were stated throughout section 1.4. was the fact that both experts and laypeople suffer from cognitive biases such as confirmation and the contextual bias. Although a great volume of research has been claiming that forensic experts suffer from contextual bias, it seemed that there were also gaps in this knowledge. Firstly, the need to identify how experts and laypeople behave towards different types of context and to observe the influences that external variables have within their decision-making process regarding their performance. Secondly the importance to differentiate how different types of contextual information affect peoples' decisions.

Before focusing the literature review provided in sections 1.1. to 1.4. to the field of forensic sciences, another topic of interest was discussed in section 1.5., namely motivation and its role in decision-making processes. In this section, the concept of motivation was described, and it was demonstrated to be an important issue for the study of decision making. Throughout the section, it was also described that

individuals can have one of two types of motivation. There are Normative Oriented people and Mastery Oriented people. This was of particular interest as Normative Oriented individuals have their motivation promoted by external sources, such as contextual information. Therefore, a question that is addressed within this thesis is related to the possibility that having contextual information can promote fingerprint examiners' motivation.

2. METHODOLOGY

2.1. Introduction

This thesis began by identifying the existing theory and literature, as well as the gaps and areas in research about decision making and cognitive biases within forensic sciences which may need additional reflection, namely when it comes to accessing contextual information. The current chapter presents the reasons why a mixed-methods approach was the most suitable choice to analyse the research questions explored in this thesis. The research questions addressed in this thesis were related to a number of aspects relevant to the decision-making processes in fingerprint analysis, namely during the verification phase. Following this methodological explanation, the design of each study, including the research questions and hypotheses are also presented. The ethical issues and limitations of the studies that were conducted will finish this chapter.

2.2. Justification for choosing a mixed-method methodology

In this thesis, the methodological approach that was chosen was a mixed-methods approach, combining complementary quantitative and qualitative data. This was a decision made as the aim of the thesis was to explore the decision-making processes in a very specific phase (the Verification) of the fingerprint analysis process (explained in the literature review as ACE-V). However, to achieve that, two types of studies were required. A set of quantitative experiments and a single qualitative study. In the first two studies, a quantitative approach was used enabling independent variables to be manipulated, in order to determine their effects on dependent variables of interest. In these studies, it was explored whether contextual information could induce cognitive biases and affect individuals' performance. This was achieved by mimicking verification tasks similar to what fingerprint examiners experience in their daily work. Since one of the previous sections in the literature review stated the importance of differentiating lay-people from experts, the set of quantitative experiments was applied to lay-people and real fingerprint examiners. Another study was only applied to fingerprint examiners in their forensic laboratories. In addition to the quantitative experiments, in the final study, a

qualitative approach was applied where semi-structured interviews were used to gather information about fingerprint examiners' experiences and attitudes, which were then analysed thematically.

These two methodological views (quantitative and qualitative) should not be considered as in opposition to one another, but rather they are complementary approaches. Whereas quantitative research mainly focuses its attention on collecting numerical data and testing hypotheses, the qualitative approach describes what is observed in real-life situations, enabling researchers to generate new hypotheses and/or theories (Johnson & Christensen, 2008). During the empirical work conducted throughout this thesis, the choice of having two types of methods promoted the convergence between two different types of data that were collected. Since one of the methods applied was a qualitative method where interviews were conducted, individuals were able to feel that they were more than mere participants carrying out a set of experimental tasks. By providing a space for their opinion throughout the interviews, it was expected that participants felt that they could actively be part of changes in the field of fingerprint analysis. This type of mechanism occurred in domains such as school improvement where students were asked to participate with their opinion in order to be active in the changes that would be implemented in their school (Lodge, 2005) Within the research on the topic of cognitive biases within fingerprint analysis, it is possible to observe several studies where a quantitative approach has been produced. However, this thesis focused its attention on an additional variable, the motivation that fingerprint examiners have when conducting their work. To access motivation as a variable, this thesis used the same methodology as a particular study that was identified, which used a qualitative approach (Charlton, Fraser-Mackenzie & Dror, 2010).

Researchers described the discussion between advocates of quantitative and qualitative data as a debate which eventually ended by both sides without an agreement or any possible point of convergence in some of the topics being discussed (e.g. data validity). After observing the limitations to converge both perspectives, researchers started to defend the possibility of using both paradigms as well as the research strategies that were used within each, paving the way for the mixed methods approach (Tashakkori & Teddlie, 1998).

Mixed methods gained its value by being an empirical approach within the continuum between the classical quantitative strategies and the emerging advances of qualitative methodologies (Creswell, 2014). There are three pillars that are essential that researchers ensure to use mixed methods as Creswell, Klassen, Clark and Smith (2010) pointed out. Firstly, researchers should focus on understanding and analysing real-life questions, multi-level perspectives and cultural influences. The second pillar recommends researchers to employ rigorous quantitative instruments to assess the magnitude and frequency of constructs that are found as well as a rigorous qualitative design to explore the meaning and understanding of constructs. Authors also stated that this type of approach should use multiple methods such as intervention trials and in-depth interviews which are intentionally combined to draw on the strengths of each, which is the approach adopted in this thesis. And finally, research should frame the investigation within philosophical and theoretical positions. Associated to these views, two paradigms can be applied to sustain theories that individually support each research method. Linked with quantitative views, it is possible to find models where the paradigms are positivist/empiricist. On the other extreme of the continuum, one finds a constructivist / phenomenological paradigm to explain a qualitative approach (Tashakkori & Teddlie, 1998).

Johnson and Christensen (2008) mentioned three actions that researchers should follow in each branch of a mixed-method approach. Within the qualitative extreme, researchers should:

- Start making some observations;
- Look for patterns within the observations; and,
- Attempt to generalize some operational reasoning from the observations.

On the other hand, the quantitative viewpoint asks researchers to:

- Organise a hypothesis [commonly based on previous research];
- Collect data; and,
- Accept or reject the formulated hypothesis.

The abovementioned framework was the most fitting framework for this thesis as the first part of data collection started with some forensic laboratories visits promoted by PhD supervisors and other project associates[4]. After a number of visits, it was possible to operate a number of questions and create hypotheses. In the end it was reasonable that some work should be considered to use a quantitative approach as well as some of the work should be a continuation of the first visits in order to seek for patterns and explore qualitative aspects of the fingerprint examiners' work. Hence the choice for the mixed-methodology for this thesis.

2.3. Research design

Greene, Caracelli and Graham (1989) presented five possible rationales for consideration when using a mixed methodology: (1) triangulation, (2) complementary, (3) development, (4) initiation and (5) expansion. The first two rationales were used in the design of this thesis.

Triangulation was used to converge, corroborate and compare results from quantitative studies with the results retrieved from the qualitative study. This was important as the quantitative studies in this thesis explored the potential sources of bias that influenced erroneous decisions during tasks that mimic a verification phase of the ACE-V either when applied to lay-people or fingerprint examiners. The qualitative study explored, amongst other factors, the motivation that fingerprint examiners have during their work. Hence, the analysis of the convergence, as well as the divergences of these two approaches, were very important for better understanding the potential of biases associated with contextual information that was included in experiments and which not only promoted erroneous decisions but also influenced examiners' motivation.

[4] The INTREPID Forensics Programme had different associates. Amongst them, some were fingerprint examiners who worked at bureaus and facilitated visits for some of the students within the programme.

The complementary rationale provided the possibility to elaborate, enhance, illustrate and clarify results from the quantitative experiments with the qualitative study that was conducted. It also allowed to elaborate in a more precise and elegant way the impact and importance of examiners' motivation to carry their work, by analysing trials' accuracy and correlating those results with the types of contextual information that were manipulated. Finally, it was also possible to verify associations with contextual information which could either be a source of motivation, a source of bias, or perhaps both. The correlation between those two variables will enable a discussion about what type of setting examiners should work in, as well as to discriminate whether the contextual information can be considered a source of motivation, even if it is a source of erroneous decisions as other studies have demonstrated before.

Results retrieved from both rationales have the potential to unlocked the link between the performance and motivation that examiners have. By gaining access to that, it was possible to discuss the factors and exercises/tests that agencies should take into consideration when recruiting new examiners to focus not only in technical exercises but also to gather subjective data such as the motivation one must have to be successful at a job like fingerprint analysis.

In this thesis, fundamental research questions were explored by testing the hypotheses that were conducted within the quantitative and qualitative studies. Since this thesis focused on the specificities of the verification phase within the ACE-V, when this thesis refers to the *work* that fingerprint examiners do, it is actually referring to tasks that mimicked a verification phase of the ACE-V process or to the procedures that fingerprint examiners carry out within that specific phase of the ACE-V process.

There were four research questions addressed in the three studies of this thesis. Hypotheses for each of the research questions are articulated in each of the empirical chapters.

1. Are fingerprint examiners influenced by contextual information to the same

extent as laypeople during a pattern recognition task? *(Chapter 3)*

2. What are the effects that different types of contextual information have on fingerprint examiners' performance? (*Chapter 3, 4 and 5*)

3. Are there significant differences in the performance, motivation and need for cognition of fingerprint examiners that have different levels of experience and who work within different conditions (e.g. methodological approach and the quality standards implemented) in their forensic bureau? *(Chapter 4)*

4. What are the factors that influence fingerprint examiners' motivation when carrying out the ACE-V process within their forensic bureau? *(Chapter 4 and 5)*

2.4. Dependent variables

Dependent variables were analysed throughout the quantitative studies. In study 1, accuracy and response time were analysed and combined with 1 independent variable which was the category of the participant (i.e. if they were experts or laypeople). Study 2 measured four dependent variables: accuracy, response time, level of motivation and need for cognition. Those four dependent variables were combined with 4 independent variables (level of experience, type of comparison, type of methodological approach and bureaus' accreditation).

Accuracy was measured by assessing the number of correct answers that participants gave in the first two quantitative studies that were conducted. Quantitative data was retrieved from experiments that attempted to mimic a verification task similar to what is carried by fingerprint examiners when carrying the last phase of the ACE-V process. Two experiments were carried with this type of setting. The first one was deployed using an online setting, whereas the second was conducted in forensic laboratories where fingerprint examiners worked.

In both quantitative experiments, two questionnaires were also applied, one that aimed to assess the level of motivation that participants had at the end of the task (WEIMS), and a second one which assessed their level of cognitive effort (NfCS). The motivation was also qualitatively assessed throughout interviews that were carried out on-site in forensic laboratories.

To answer the research questions described above, the quantitative and qualitative results were independently analysed and then combined to determine whether they would converge by using the two rationales that were described above (triangulation and complementary). Quantitative studies allowed the researcher to understand the effects of contextual information could be within the ACE-V and whether this type of stimuli/condition could be a source of bias that affects performance. On the other hand, the qualitative workstream focused on the motivation that fingerprint examiners have during their caseload, including the motivation that they have due to contextual information. This was extremely interesting to observe and see potential links between a potential source of bias and motivation at the very same time.

2.5. Quantitative studies (Chapters 3 and 4)

The online experiment (Chapter 3) aimed to analyse differences between accuracy and motivation of laypeople and experts in the field of fingerprint recognition when both groups carried out a task of pattern recognition that had two types of stimuli – artificial fingerprints and excerpts of text. This type of study also known as a web-experiment or internet-based experiment research study (Hewson, Yule, Laurent & Vogel, 2003), is accessed by using internet services that take participants to an online platform (Reips, 2007) similar to the one that has been used in this thesis – Qualtrics (2019).

Internet-based experiments have been widely used when carrying out research within the field of decision making (Wald, Gray & Eatough, 2019) in contrast with other psychological fields such as neuropsychology or even perception (Reips, 2000). A point of interest within the use of internet-based experiments is the reliability one can have from the outcomes of such a research design. Gosling, Vazire, Srivastava and John (2004) mentioned that internet sampling can contribute to many areas of psychology comparably to traditional methods. The authors proposed two arguments to support the use of internet-based data collection. Firstly, the difference of sample size they observed in a variety of studies that used either a physical data collection or an internet-based data collection. The former method of collecting data had a significantly smaller number of participants (n = 510) that the

latter (n = 361,703) where sampling was retrieved from the internet. Secondly, the authors observed that internet-based experiments were not adversely affected by nonserious or repeated responders, making the results retrieved from online methods consistent with traditional methods of data sampling. Additionally, in terms of generalization, internet sampling derives findings that are rather diverse in terms of gender, socioeconomic status, geographic region and age. The use of internet-based experiments promotes the prevention of response bias from participants in the experiments as well as not disrupting the type of sampling since internet samples also provide quality data for research (Hewson, 2008).

Concerns were also raised regarding some disadvantages of using internet-based experiments. Reips (2000) presented four disadvantages:

(1) Multiple submissions – the same participant submits more than one set of answers;

(2) lack of experimental control – researchers not having total control of the experimental environment;

(3) self-selection – participants' entries are biased due to the type and topic of the experiment;

(4) drop out – participants dropping out the experiment without the control of the researcher.

The internet-based experiment that was conducted in this thesis had some limitations regarding, for instance, the environment that participants were in when carrying out the tasks within the experiment. It was not possible to control the environment where participants carried out the experiment nor the type of devices they used (i.e. the size of the screen, type of light – which are of particular importance for a pattern recognition task). However, this thesis aimed to mitigate some of the concerns raised by Reips (2000) using suggestions from research (Wald, Gray & Eatough, 2019) advising experimenters on the use of internet sampling. Concerning the multiple submissions, participants were informed that their participation should be serious in order to retrieve quality data and that they

should not submit more than one set of answers. They were also informed that their IP address was recorded to prevent multiple submissions.

To increase the randomization of participants (instead of having participants only linked to the forensics field), participation in this experiment was advertised throughout a range of channels such as (i) personal contacts of the principal researcher, (ii) mailing lists that were accessible, (iii) social media, (iv) through the university newsletter and finally (v) through mailing lists that the first supervisor of this thesis was able to provide.

A prize draw was also one of the three possible solutions to prevent drop-outs since individuals were only enrolled in this prize draw if they finished the experiment. To prevent participants from providing random answers only to get enrolled within the prize draw a first analysis of the scores was assessed to validate missing answers or entries that seemed invalid. This was combined with a warm-up phase as well as with the possibility of having the experiment available in three different languages (English, French and Portuguese).

Still related to using the internet as a means to carry out experiments, more tools are currently possible to include besides only simple stimuli integrated into video, audio and/or images (Cohen, Manion & Morrison, 2007). In the past, samples retrieved from the internet were not considered as representative of the population as they should be, due to people's accessibility to the internet. However, this view has changed (Gosling & Mason, 2015), and internet-based experiments are now considered to be a beneficial and accurate way to observe a range of experimental phenomena. Having considered the positive and negative aspects of the online experimental approaches, this thesis considers it to be a valid method for the research questions of interest.

The second study (Chapter 4) was conducted on-site in forensic laboratories that the researcher physically visited to meet with fingerprint examiners. Within this study, the experiment was set up using OpenSesame software (Mathôt, Schreij, Theeuwes, 2012) and by using python programming (van Rossum, 1995). Throughout the visits, it was possible to observe the physical spaces where fingerprint examiners worked at the time of the visit, discuss some aspects of the

work carried out in the laboratories and to engage with potential participants. The visits were all facilitated by gatekeepers who were previously contacted by the researcher. Arrangements for ethical/clearance approvals within the bureaus were met previous to the visits to happen. In some laboratories, once the data collection was finished, it was also possible to deliver some training related to methods of prevention of cognitive bias or to carry out other tasks such as recruitment and selection of new hires.

This experimental study was conducted with 14 forensic bureaus from 9 countries. The sample only included fingerprint examiners who volunteered in each laboratory that was visited. Using fingerprint examiners as participants was extremely important in order to achieve higher levels of realism to support generalizations from the results that were observed. All of the participants' responses were anonymised and there were no incentives or benefits to motivate participation. In addition, there were no consequences for those who did not want to participate. In terms of control, this design was assumed to be more controlled than the online study for a number of reasons. The presence of the researcher in the room where the experiment was being conducted allowed the researcher to ensure that conditions were always similar between trials and even between laboratories. A room in which to conduct the study was always requested in order to control conditions such as light and noise. The screen and software setting that was used (in this case the laptop of the researcher) was also the same for all participants. Other people were not present in the time of the studies except for participants located at the Chinese National Police who were accompanied by a translator, even though the information/guidelines that were provided during the experiment were available in four different languages (English, French, Portuguese and also in Chinese).

2.6. Materials

2.6.1. Artificial Fingerprint Dataset

The artificial fingerprint dataset used as stimuli was created by using sFinge software (Cappeli, 2015). This open-source software allows the creation of artificial fingerprint images. Four features could be manipulated within the software, the

direction of the ridges, the density of the ridges, the type of pattern (whorl, loop or arch) and also the noise that each image has.

To have a ground truth database rational, artificial fingerprints were created in pairs. A first image (original fingerprint image) was created without any noise and with good quality aiming to mimic what fingerprint examiners observe when they use ten-print cards or fingerprints retrieved by using controlled collection techniques (either physical or digital), and a second image (distractor fingerprint image) which was created by applying noise and distractor features to the first image. By doing this, it was possible to know what the correct answer of matching pairs with 100% certainty was.

Within study 1 and 2, there were three types of pairs that participants needed to compare within trials: matches, non-matches and close non-matches. Matching pairs had the same pattern and ridge flow and differed on the noise presented in the distractor fingerprint image. Non-matching pairs were equally created but those differed on their pattern and ridge flow. Close non-matching required more work to be created. After choosing the original fingerprint images to be used for those pairs, besides applying noise to create the distractor fingerprint image a subtle manipulation of the ridge details was applied, i.e., close non-matching pairs matched in the pattern they had, however, they were not matching as their ridge flow were subtly different.

2.6.2. Excerpts of Text Dataset

Excerpts of text were used to compare the performance between laypeople and fingerprint examiners in the pattern recognition task that is presented in Chapter 3. To have stimuli that was familiar to laypeople the same way artificial fingerprints were to fingerprint examiners, excerpts of text were used based on the research that previously compared laypeople with forensic experts in other fields such as Pattern Visual Recognition (Nagy & Zou, 2002), Forensic Signature Analysis (Dyer, Found & Rogers, 2006), Speaking Recognition (Alexander, Botti, Dessimoz & Drygajlo, 2004) or Forensic Handwriting (Pervouchine & Leedham, 2007).

All excerpts of text were retrieved from academic journals with the permission of the authors to be used for this specific research experiment. Equally to the manipulation of artificial fingerprints, each excerpt of text was manipulated using design software to add noise (i.e. dark spots and blurred areas) and disturbance (i.e. typos, different words, punctuation) to the image and to transform some of the pairs into non-matches.

2.6.3. Stimuli Development – Contextual information

One of the main objectives of the experimental studies was to observe if contextual information could be a source of cognitive bias and therefore affect accuracy within a verification task. In study 1, participants were asked to carry out a pattern recognition task using fingerprints as well as pieces of text. For that reason, contextual information that was developed was relevant to either the fingerprint trials or the pieces of text trials. Study 2 was conducted only with fingerprint examiners and therefore only used the contextual information related to the fingerprint task and its features.

The justification for the choice of this type of contextual information was related to previous where authors observed the effects of contextual information within forensic professionals' performance. As an example, Dror and Hampikian (2011) described in their study that DNA examiners spent more time on the tasks when the context described a serious crime (e.g. sexual assault).

In each experiment, four blocks of trials were presented. Three of these had manipulated contextual information and a fourth block did not have any contextual information (control). For both types of stimuli (artificial fingerprints and excerpts of text), the manipulation of contextual information (dependent variable) was arranged accordingly to fit in the theme of the stimuli regarding (1) its origin, (2) previous conclusions reached by a senior third party with expertise in the field and (3) background associated to someone linked with the stimuli which was being presented. Each type of contextual information had two possible manipulations as it is shown in the table below.

Blocks	Excerpts of text		Fingerprints	
Source of the stimuli	Undergraduate assignment	journal articles	Volume crime	Major crime
Senior third-party previous conclusion	Correct previous conclusion	Incorrect previous conclusion	Correct previous conclusion	Incorrect previous conclusion
Background of the stimuli	The author is an undergraduate student	The author is a lecturer	Suspect with a criminal record	Suspect without a criminal record
Control	No context			

Table 7 – Type of contextual information within empirical studies (Chapter 3 and 4)

2.6.4. Work Extrinsic and Intrinsic Motivation Scale (WEIMS)

The WEIMS (Tremblay, Blanchard, Taylor, Pelletier, Villeneuve, 2009) (appendix C) has been widely used to assess extrinsic and intrinsic motivation within organisations. The scale has 18 items, that require participants to answer each item using a Likert-scale. In each item, participants face a statement in which they assess the extent to which it corresponds to the reasons why they are presently involved in the work they carry out on a daily basis. Positive scores on the scale indicate an internally motivated profile whereas negative scores suggest externally-driven individuals regarding their motivation.

The WEIMS was used to observe whether there were significant differences regarding the level of motivation of fingerprint examiners from different rankings of experience. The rationale in using this scale and the association with Hoffman (2015) was built to observe whether the suggestion that there are individuals who are motivated by internal inputs, such as the specific task of a pattern recognition within a fingerprint analysis and others who are externally motivated perhaps by knowing more than the strictly provided by the task of analysing a fingerprint (i.e. for instance by the context of the crime) could be reasonable.

2.6.5. Need for Cognition (NfC)

To assess the cognitive effort that individuals had in the experimental studies, the Need for Cognition Scale – short version (Cacioppo, Petty & Kao, 1984) was applied to all participants (appendix B).

The short version of the NfC is an 18 item scale that has been used to assess the level of cognition that individuals show in a specific task, (i.e. the level of enjoyment of applying their cognition to a specific task) (Cacioppo, Petty & Kao, 1984). Gomes, Santos, Gonçalves, Orgambidez-Ramos & Giger (2013) argued that individuals with higher levels of need for cognition demonstrate higher intrinsic motivation that is related to the process of thinking about very specific tasks such as solving puzzles or riddles. On the other hand, people with low scores on the NfC scale are less prone to focus on central elements of tasks and usually take peripheral views as well as rely on heuristics to solve challenges.

2.7. Qualitative study (Chapter 5)

For the success of the objectives proposed in the research questions within the qualitative workstream, this thesis conducted on-site interviews with fingerprint examiners. Interviews can have different formats regarding their structure such as semi-structured, unstructured or in-depth (Mason, 1994).

An unstructured interview will be extremely open and flexible. The agenda that is used to achieve the research goals can take a diversity of directions, and although this type of data collection can provide the research with precious information, it can take a very long time period (Harrel & Bradley, 2009). Unstructured interviews are more comparable to a conversation rather than the definition of the interview itself (Jamshed, 2014). On the other extreme, one finds the structured interviews, which are the most controlled type of interviewing, as the questions are fixed and asked in a specific order. This type of data collection does not allow much flexibility towards the respondents' thoughts and its direction.

For the qualitative data that was collected within this thesis, a semi-structured interview protocol was chosen to be the most suitable approach to be followed. According to Harrel and Bradley (2009), semi-structured interviews allow the researcher to gather opinions, perceptions and attitudes as well as experts' knowledge, and the description of the processes that are carried out in case-work. One of the main objectives of the qualitative study was in fact, to address open-ended questions and let fingerprint examiners provide their viewpoints on topics such as the effects of contextual information and the work carried out within a fingerprint bureau.

The interviews took an average time of duration according to what Dicicco-Bloom & Crabtree (2006) suggested (± 1 hour). Commonly semi-structured interviews follow a list of questions that were previously designed and piloted with a small sample aiming to observe if these items cover the topics the research wants to explore (Creswell, 2007). In this study, questions used in the interviews were discussed with a pool of experienced fingerprint examiners from different laboratories in different countries and some adjustments were made to the final list of questions (appendix A). These discussions started with a set of questions and observations made by the researcher following a number of bureaus visits. Questions and observations were related to the day-to-day work that fingerprint examiners carry within the bureau plus some of the evidence that research brought to light in the field. After those discussions, a set of questions started to be planned with research supervisors and with examiners in order to validate the importance of those. Questions from the list, consider a number of variables such as the type of interview that the researcher wants to conduct, the interviewee and the language that should be applied (Fylan, 2005).

According to Willis (2005), during interviews, researchers can make use of 6 types of cognitive inquiring strategies to apply within the interview's probing questions and retrieve higher quality responses: (i) comprehension/interpretation, (ii) paraphrasing, (iii) confidence judgment, (iv) recall, (v) specific and (vi) general. These cognitive inquiring strategies have one of two different origins, either proactive or reactive probing. While proactive probes encourage interviewees to carry/develop the reasoning within the order of the design that the researcher has for the interview, reactive probes mark a specific moment that the researcher wants

to explore further or ask the participant to reflect on (Willis & Artino, 2013). Throughout the interviews that were conducted in this study, four types of these strategies were used. The comprehension/interpretation was used to better explore some terms that respondents mentioned during their answers by including a cognitive inquiring such as "What does the term 'formal educational program' mean to you?". The recall was used to analyse individuals' rationales (e.g. "How did you come up with your answer?"). Specific strategy cognitive inquiring was used to analyse details within answers (e.g. "Why do you say that you think it is very important that examiners participate in continuing training?"). Finally, general strategies of cognitive inquiring were used in distinct moments such as when individuals hesitated by noting those moments using for instance "I noticed that you hesitated. Tell me what you were thinking".

Because semi-structured interviews can encourage respondents to share very sensitive information within certain topics the researcher wants to cover, a very explicit ethical plan should be given previously the interview starts (Fylan, 2005). Privacy and confidentiality when carrying out individual interviews are two concepts that one can never ignore. This was addressed at the beginning of every interview, as well as the notice that participants could stop the interview at any time or even withdraw from the study with any kind of consequence.

As mentioned above, this thesis used two techniques of a mixed-method approach: triangulation and complementary. The inputs retrieved from respondents' answers in the qualitative study were extremely helpful to understand some of the effects that were found in the experimental studies such as the impact of contextual information on accuracy and/or motivation. For that to happen, questions were introduced with a certain degree of flexibility due to the versatility of this mean to collect data (Fylan, 2005), meaning that the order that these were asked during the interview was not always the same with every participant.

Some of the disadvantages of conducting semi-structured interviews mentioned by Adams (2015) are related to the time taken that this design asks for as it should be considered not only the time of the interview but also the time to transcribe the interview, the access to participants and the number of participants. All of these

issues were considered in the design of this thesis. Firstly, the time of conducting the interviews was balanced by the use of software to help to transcribe the interviews. The access to participants was facilitated by the contacts that were made with gatekeepers who openly invited the researcher to visit forensic laboratories and recruit volunteers for the study. As it was demonstrated in the results of the qualitative study (Chapter 5), there were two very distinct cognitive approaches regarding the work that a fingerprint examiner does which were then complemented by the results of the quantitative studies as described at the beginning of this chapter regarding the use of mixed-methods.

2.8. Ethical concerns

All research conducted in this thesis received ethical approval from the University of Leicester's Ethics Committee. Two ethics applications were submitted covering (i) the quantitative studies, and (ii) the qualitative study. Since the on-site quantitative study and the qualitative study were conducted within forensic bureaus premises, at some of the bureaus, police vetting and visiting applications needed to be submitted by the researcher.

Informed consent from all participants was obtained previous to any experimental work, guaranteeing what Bryman (2012) and Bulmer (2008) mentioned regarding the need to ensure confidentiality and protection of all data collected. For the quantitative studies, an explanation was provided regarding the objectives of this thesis and the expected outcomes. This explanation was provided by text, either on the screen (online experiment) or in writing (on-site experiment) previous to the experiment starting as well as information regarding the ethical approval, the details of the researcher and supervisors and all of the objectives of the study. In the online experiment, participants needed to press the "I agree" button to provide consent and proceed to the experiment. In the on-site experiment, participants needed to sign the informed consent form previous to starting the experiment (appendix D).

Since the quantitative studies analysed accuracy, participants were made aware that a score would be allocated to each participant. This could have been an obstacle in the on-site experiments as individuals may have thought that their superiors would

have access to their answers and scorings. However, all participants were assured that responses would be anonymised and coded into a dataset that could not link individuals to their scores. They were also assured that only the researcher would have access to the raw data.

For the qualitative study, an informed consent form was also provided (appendix F), where the objectives and details of the study were described as well as the researcher and supervisors' contact information. The interviews were not expected to include sensitive issues except for the question "what motivates you to do your job?", which was a broad and open question. In one particular situation, an individual cried during the interview and some others presented emotional reactions. Interviews that get emotional can cause harm to the interviewees (Elmir, Schmied, Jackson & Wilkes, 2011). Even though this was not expected to happen, the researcher was prepared to respond to this by using his knowledge and experience in practising clinical psychology and counselling and promoting a comfortable environment where empathy, appropriate timing and rapport was promoted. Participants were asked if the interviews could be recorded and explained that all recordings would be saved securely and that only the researcher would have access to it. Confidentiality was also ensured for all participants. Except for one participant, all individuals accepted the interview to be recorded, and the researcher also took personal notes during the interviews.

2.9. Conclusion

This chapter outlined the research design that was applied within the studies conducted. Three overarching research questions were developed, which were related to a number of aspects relevant to the tasks that fingerprint examiners carry out within their workload. To answer those questions, 3 studies (2 quantitative and 1 qualitative) were conducted. Due to the nature of the research questions, a mixed-method methodology was found to be the most suitable research approach.

Mixed methods allowed the researcher to gather information from the quantitative and qualitative sources and to create connections between both types of data by implementing triangulation and complementary. Besides the justification of the

methodology's rationale that was applied, this chapter also described the materials that were used and how these link with the topics that were studied, such as the effects of contextual bias and its effects on accuracy, motivation and cognitive effort.

Procedures were described in each study conducted, including how the recruitment of participants was achieved and the implementation of the studies in three different settings, namely (i) online, (ii) on-site experiment and (iii) on-site interviewing. Finally, the ethical considerations were also acknowledged across all three studies. Further details of the methodology specific to each study are included in the following empirical chapters.

3. EFFECTS OF COGNITIVE BIASES ON PERFORMANCE OF FINGERPRINT EXPERTS AND LAYPEOPLE DURING PATTERN RECOGNITION TASKS

3.1 Abstract

A substantial amount of literature has challenged the objectivity of fingerprint examinations in recent years. The focus of some research has been the role of cognitive bias within the decision-making process of fingerprint experts. Even though research already found that fingerprint examiners perform better than laypeople in a fingerprint comparison task, studies have not addressed the differences between experts and laypeople regarding the influences of different types of contextual information on their performance. This study addressed this gap by applying to a group of experts (n = 41) and of laypeople (n = 57) two different types of pattern recognition tasks. Results suggest that laypeople are similarly influenced by the same type of bias than experts when stimuli are familiar to them. It was also observed that some types of contextual information did not influence both types of participants' performance.

3.2 Introduction

People tend to frame decision-making tasks under two conditions, (i) the characteristics of the guidelines and/or the habits they use to decide and (ii) the evaluation of the problem they need to decide upon (Kahnemann & Tversky, 1981). To complete the set of procedures one needs to follow to decide, individuals apply different strategies that are constrained by variables such as the context (Kerstholt & Raaijmakers, 1997). There are different strategies that one can select to reach a decision, however, people tend to use short-cuts to make decisions (Payne, Bettman and Johnson, (1993) in order to achieve accurate performance, while simultaneously making minimal investment (Kerstholt and Raaijmakers, 1997), and conserve their cognitive effort (Aronson, 1999). Yet, these shortcuts that individuals make use of are also potential sources of bias such as the confirmation bias (Kahnemann & Tversky, 1973), or contextual bias (Kassin, Dror & Kukucka, 2012) since human reasoning is often based on heuristics (Pi, Parisi & Luppi, 2014).

One of the frameworks that proposes how to assess expertise is the Performance-Based Approach which suggests ten core characteristics for an individual to be identified as an expert (Shanteau, 1988). Amongst the characteristics, experts are expected to conduct a number of tasks better than laypeople such as extracting information, communicating effectively, and developing innovative strategies. Experts' characteristics related to their judgment are expected to develop within a *relativistic* perspective (Weiss & Shanteau, 2014), i.e. individuals are expected to develop their judgement skills within a continuum until they achieve high levels of expertise. Regarding this core skill – the judgement – Weiss and Shanteau (2003) suggested expertise to be observed in four different types: (i) based in a qualitative and quantitative evaluation, (ii) based in a projection, i.e. when an expert provides a certain prediction, (iii) based in the communication and capacity of providing instructions and (iv) based in the execution itself.

Research in fields such as medicine (Saposnik, Redelmeier, Ruff & Tobler, 2016; Pines, 2006), aviation (Gilbey & Hill, 2012) or fingerprint analysis (Kassin, Dror & Kukucka, 2013), where expertise development is essential, has challenged the idea that experts are immune to cognitive biases as discussed by Pi, Parisi and Luppi (2014) regarding the effects of cognitive biases affecting either laypeople or experts. Tversky and Kahnemann (1974) suggested that intuition based on three heuristics (availability, representativeness and anchoring) could be the primary process for individuals to experience biases such as the confirmation bias (Nickerson, 1998) or contextual bias (Todorović, 2010).

In Chapter 1 of this thesis, forensic experts were presented as individuals that are susceptible to potential biases. In the different fields the chapter described such as bite marks (Osborne, Wood, Kieser and Zajac, 2014), DNA (Alexander, 2015a; Alexander, 2015b; Dror & Hampikian, 2011; Garret & Neufeld, 2009) or fingerprint comparisons (Tangen, Thompson & McCarthy, 2011; Kassin, Dror & Kukucka, 2013), experts observed their performances lacking quality in specific situations due to sources of errors related to either confirmation or contextual bias. To evaluate whether laypeople and experts can be affected by these two types of cognitive bias during a pattern recognition task, this study created a set of tasks where laypeople and forensic experts were asked if two stimuli were a match or a

non-match. In order to address the representativeness of the stimuli, two types of stimuli were used, a set of artificial fingerprints that fingerprint experts could relate to regarding their daily work and excerpts of texts that laypeople could relate to as well.

3.3 Method

3.3.1 Sample

This study had two types of participants; One sample of fingerprint experts, and one sample of laypeople. Individuals were considered experts if they had been working as fingerprint examiners, whereas individuals that did not work as fingerprint examiners were considered laypeople.

Individuals from both groups were selected through a snowball method where a hyperlink generated by the software used to run the experiment (Qualtrics) was disseminated through social media, mailing lists as well as direct contacts with gatekeepers at fingerprint bureaus. Participants were also invited to share the hyperlink with other contacts to increase participation.

A total of 98 participants completed the experiment, 57 laypeople (M_{age} = 31.08 years, SD = 10.89, min = 20, max = 67) and 41 fingerprint examiners (M_{age} = 35.48 years, SD = 6.98, min = 23, max = 50). Fingerprint examiners' experience ranged between 1 and 23 years ($M_{experience}$ = 9, SD = 5.35, min = 1, max = 23). Individuals within the expert group needed to be employed as a fingerprint examiner at the time they participated the experiment to be considered as a participant.

3.3.2 Design

An ethics application was approved by the University of Leicester Ethics Committee in order to carry out the online experiment. There were no expected harms for participants. The objectives of the study were explained at the beginning of the study, before starting the experiment as well as the request to tick a box providing consent to participate. Contact details of the researcher and his supervisors were also provided in case any clarification was needed or also for

requests to withdraw from the study. Participants were told that there would be no adverse consequences if they wished to be removed from the study for any reason.

Two conditions differing in the type of stimuli were conducted. Within each condition, a pattern recognition task was presented consisting of either artificially generated fingerprints or excerpts of text. As mentioned in Chapter 1, people who are described as experts in a certain field do so due to the developed experience they have within the field itself (Weiss & Shanteau, 2014). Fingerprint examiners are considered experts in fingerprint comparisons as they have greater experience working with that kind of stimuli. Aiming to compare the performance of laypeople with fingerprint examiners in a pattern recognition task, the excerpts of text were used to mimic stimuli that were also familiar to laypeople. This design built on the research that previously compared laypeople with forensic experts in other fields such as Pattern Visual Recognition (Nagy & Zou, 2002), Forensic Signature Analysis (Dyer, Found & Rogers, 2006), Speaking Recognition (Alexander, Botti, Dessimoz & Drygajlo, 2004) or Forensic Handwriting (Pervouchine & Leedham, 2007). To prevent participants from influences associated with the topics of the excerpts, all pieces of text were retrieved from journal articles with the same topic, in this case, gamification in youths.

In each condition, four blocks with two trials each were presented to participants. In addition to a block used as the control, each block within each condition was assigned one type of contextual information. There were three types of contextual information: (i) information related to the source of the stimuli, (ii) information related to the background of the suspected author of the stimuli and (iii) information related to a previous conclusion reached by another person for the same task. During each trial, participants were presented with two stimuli and asked to decide whether the stimuli matched or did not match. Blocks and trials in the experiment were randomly assigned to participants to prevent selection bias as suggested by Viera and Bangdiwala (2007) suggested. In the experiment, each participant completed 16 trials (2 trials per block). In total, 1,568 comparison trials were carried out in this study.

In each condition, contextual information was presented in a way that was relevant to the stimuli type in each condition. The table below summarises the contextual

information that was presented in each block within the two conditions (artificial fingerprints and excerpts of text).

Blocks	Fingerprints		Excerpts of text	
Source of the stimuli	Volume crime	Major crime	Assignment	Journal article
Background information	Criminal record of suspect	Suspect has no criminal record	Student author	Lecturer author
Previous conclusion	Correct	Incorrect	Correct	Incorrect

Table 8 – Types of contextual information included within blocks in each condition

Participants had no time limit to view slides with contextual information, however, when presented with the stimuli, they had 30 seconds to decide within a forced decision-making approach where the options were (1) match or (2) non-match. This forced decision making was implemented since tasks that ask individuals to assess the matching level of stimuli involve a side-by-side comparison, which relies less on memory (Thompson, Tangen & McCarthy, 2014) as well as the fact that having an inconclusive option (i.e. an option that is neither a match nor a non-match) would be less efficient as the value of the term "inconclusive" has been discussed in previous literature (Biedermann, Bozza, Taroni & Vuille, 2019; Dror & Langenburg, 2019). For those reasons, this study used a forced decision-making approach, which was also used in previous literature in the field of fingerprint analysis (Thompson, Tangen & McCarthy, 2014).

Before starting the experiment, participants were presented with a warm-up phase with 4 trials where two stimuli of each type (excerpts of text and artificial fingerprints) were presented. Warm-up trials were not presented with a time limit or contextual information. After the warm-up phase participants started the experiment. After completing the first two blocks, they had a chance to stop and take a break to rest before completing the last two blocks and being presented with the final closing slide. At the end of the experiment laypeople, participants could provide their contact email to be entered within a prize draw for an Amazon voucher worth £25. This prize draw was not available to fingerprint examiner participants

since these were expected to be already motivated to carry out the study's experiment. Data analysis prevented to include participations that answered randomly just to enter the prize draw by observing any anchoring patterns or randomized answers as well as the total time response each participant had and seek for outliers.

It was expected that accuracy would be affected by blocks where contextual information was present, independent of the type of stimuli. It was also expected that fingerprint examiners when carrying out the trials with artificial fingerprints, would perform better than lay-people. Similar accuracy was expected for both groups within trials where excerpts of text were used as stimuli.

Hypothesis testing explored the performance of each group regarding the level of expertise concerning the stimuli used for both types of participants. The performance was measured by an analysis of accuracy and response time for each group throughout the trials in each of the four blocks. Hypotheses in this study also explored the differences within groups and between groups regarding performance in each type of contextual information.

Hypotheses

- H1: *When comparing artificial fingerprints, fingerprint examiners will perform better than laypeople, i.e. significantly more accurate and faster.*
- H2: *There would be no significant differences in accuracy and response time between both groups of participants when the stimuli are excerpts of text.*
- H3: *Both groups of participants will have poorer performances in blocks where contextual information is included before the stimuli, i.e., they will have lower accuracy scores and take more time to complete the task*.

3.3.3 Materials

Participants completed the experiment using internet-based software made available through open-source software (Qualtrics, Provo, UT).

Excerpts of text were retrieved from academic journals with the permission of the authors to be used for this specific research experiment. After selection of the excerpts, each paragraph was manipulated using design software to add noise and

disturbance to the image and to transform images that were designed to be non-matches.

Artificial fingerprints were generated using SFinGe (Cappelli, 2015). The set of artificial fingerprints consisted of simulations of fingermarks that are recovered from crime scenes and simulations of scanned fingerprints of potential candidates that can be retrieved from automated systems such as AFIS. Within trials that presented artificial fingerprints as stimuli, there were three types of fingerprint patterns. These were presented in different proportions according to the literature that was described in Chapter 1 regarding the prevalence of those pattern in the general population (Smith and Bond, 2015): 65% loops, 35% whorls and 5% arches.

Stimuli were generated in pairs within two possible options, identical (match) or non-identical (non-match). Half of the stimuli in both conditions matched. Contextual information was created for this experiment by the author or by using news and images from open source websites.

Figure 12. Example of a match trial (artificial fingerprints)

Figure 13. Example of a non-match trial (artificial fingerprints)

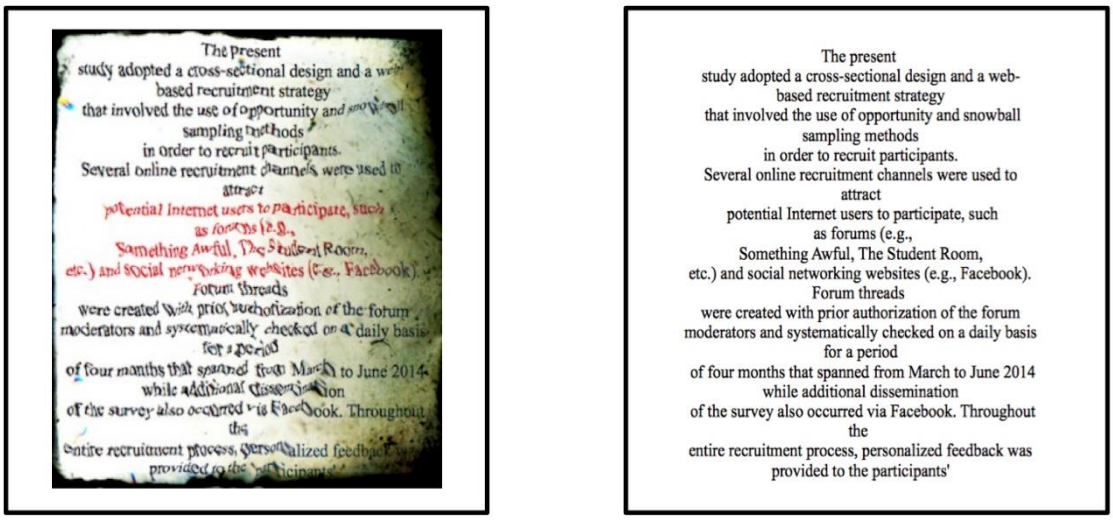

Figure 14. Example of a match trial (excerpt of text)

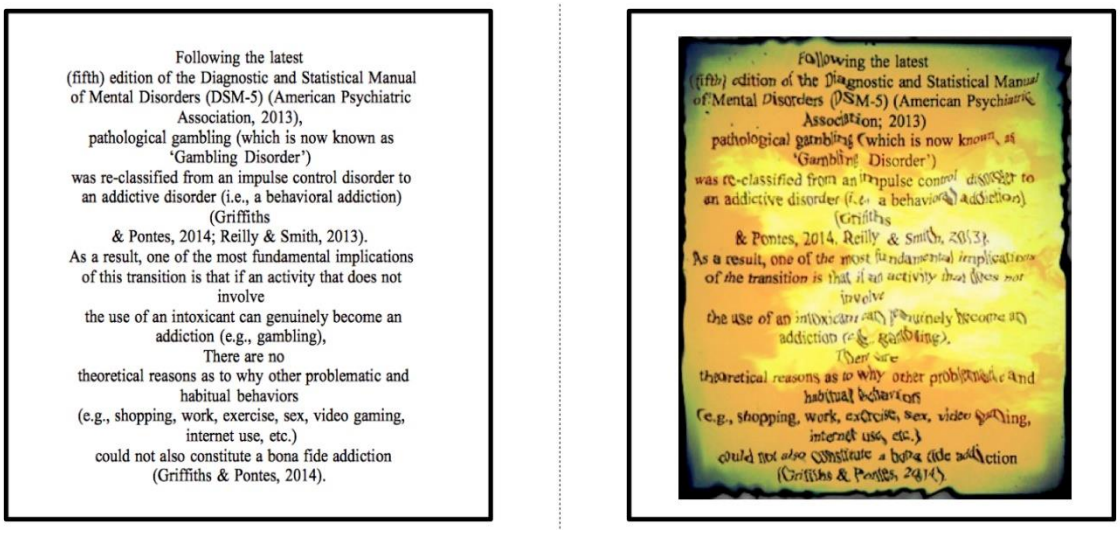

Figure 15. Example of a non-match trial (excerpt of text)

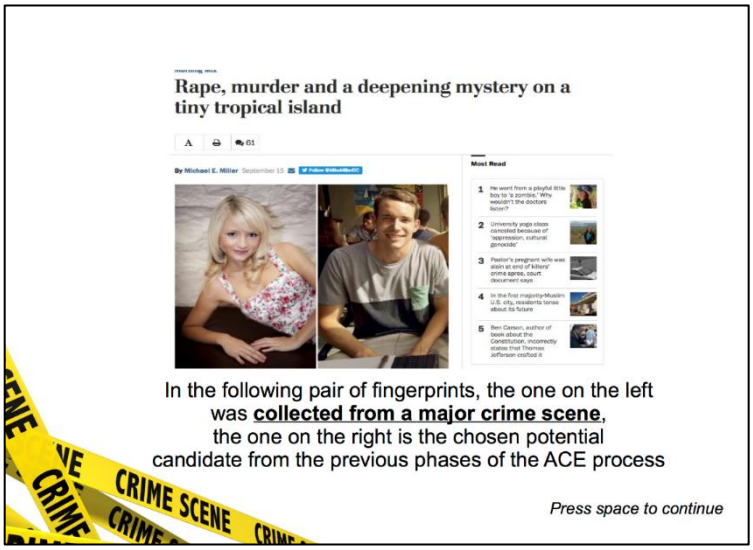

Figure 16. Example of contextual information regarding 'type of crime'

Psychology and Development

Tom Mitchell

BA Course Essay 2016-2017

Figure 17. Example of contextual information regarding types of publication

Before running the experiment, a set of the artificial fingerprints was sent to five senior fingerprint examiners (all > 7 years of experience) to confirm ecological validity (Shadish, Cook & Campbell, 2002) regarding how realistic the images were. All the senior examiners were certified and members of scientific entities such as the International Association for Identification (IAI) and members of scientific networks (ex. European Network of Forensic Services Institutes, ENFSI). The first set that was sent was not deemed satisfactory by the pool of experts and due to this, a second set was made, being unanimously approved by all five experts in the pool as being good for the practice of research within the field of fingerprints. Changes between both datasets were mainly related to the appearance of the stimuli. The second set was described to be approximated to real fingermarks retrieved from a crime scene whereas the first dataset had comments on that type of validation.

3.4 Results

Tests of normality were conducted for the outcomes observed within answers from both populations of participants (laypeople and experts). Assumptions for normality and homogeneity of variance within the distribution of the data were observed by conducting a Shapiro-Wilk test of normality ($p > .05$) as suggested by Ghasemi and Zahediasl (2012) as the sample of fingerprint examiners had less than 50 participants ($n = 41$) and the sample of laypeople was near the threshold of 50 individuals ($n = 57$).

Hypothesis 1 (H1): When comparing artificial fingerprints, fingerprint examiners will perform better than laypeople, i.e. significantly more accurate and faster

To test this hypothesis a *t-test* was conducted to investigate whether there were significant differences between experts and laypeople regarding their accuracy and response time during the artificial fingerprints condition (tables 9, 10, 11 and 12). For accuracy within the artificial fingerprints trials, results found that fingerprints examiners performed significantly better (5.54 ± 1.61 correct answers) than laypeople (4.67 ± 1.37 correct answers), $t(96)=2.880$, $p=.005$. For response time, results showed statistically significant differences, where fingerprint examiners took longer to make their decisions (163.57 ± 28.81 seconds) when compared with laypeople (137.04 ± 39.76 seconds), $t(96)=3.638$, $p=.000$. For those reasons, H1 was confirmed regarding the accuracy, however, rejected regarding response time.

	Type	N	Mean	Std. Deviation	Std. Error Mean
Accuracy artificial fingerprints	Experts	41	**5.5366**	**1.61396**	.25206
	Laypersons	57	**4.6667**	**1.36713**	.18108

Table 9 – Mean accuracy of Experts and Laypeople during trials with artificial fingerprints

		F	Sig.	t	df	Sig. (2-tailed)	SE	95% Confidence Interval of the Difference	
								Low	Upper
Accuracy artificial fingerprints	Equal variances assumed	.486	.487	2.880	96	**.005**	.30205	.270	1.469
	Equal variances not assumed			2.803	77.24	.006	.31036	.252	1.488

Table 10 – T-test of difference of mean accuracy between Experts and Laypeople during trials with artificial fingerprints ($p < .05$)

	Type	N	Mean	Std. Deviation	Std. Error Mean
RT artificial fingerprints	Experts	41	**163.5689**	**28.8158**	4.50028
	Laypersons	57	**137.0407**	**39.7567**	5.26592

Table 11 – Mean response time of Experts and Laypeople during trials with artificial fingerprints

		F	Sig.	t	df	Sig. (2-tailed)	SE	95% Confidence Interval of the Difference	
								Low	Upper
Accuracy artificial fingerprints	Equal variances assumed	4.609	.034	3.638	96	**.000**	26.52816	7.292	12.054
	Equal variances not assumed			3.830	95.99	.000	26.52816	6.927	12.778

Table 12 – T-test of difference of mean response time between Experts and Laypeople during trials with artificial fingerprints
($p < .05$)

Hypothesis 2: There will be no significant differences in accuracy and response time (RT) between both groups of participants when the stimuli are excerpts of text

This hypothesis examined the accuracy and response time of participants from both populations during the trials where they compared excerpts of the text as stimuli. Tables 13, 14, 15 and 16 described the statistics of *t-tests* for accuracy and RT. Tests demonstrated that there were no statistically significant differences in accuracy between fingerprint experts (4.78 ± 1.24 hits) and laypeople (4.38 ± 1.71 hits), $t(96)=1.254$, $p=.213$. Regarding response time for this type of stimuli, similarly to H1, fingerprint experts were significantly slower (165.85 ± 29.25 seconds) when compared to laypeople (140.16 ± 40.66), $t(96)=3.452$, $p=.001$. Hence, criteria to accept H2 were met only for accuracy but rejected for response time.

	Type	N	Mean	Std. Deviation	Std. Error Mean
Accuracy excerpts of text	Experts	41	**4.7805**	**1.23516**	.19290
	Laypersons	57	**4.3860**	**1.71916**	.22771

Table 13 – Mean accuracy of Experts and Laypeople during trials with excerpts of text

		F	Sig.	t	df	Sig. (2-tailed)	SE	95% Confidence Interval of the Difference	
								Low	Upper
Accuracy artificial fingerprints	Equal variances assumed	5.274	.024	1.254	96	**.213**	.39452	.315	-.2298
	Equal variances not assumed			1.322	95.99	.189	.39452	.298	-.1978

Table 14 – T-test of difference of mean accuracy between Experts and Laypeople during trials with excerpts of text
($p < .05$)

	Type	N	Mean	Std. Deviation	Std. Error Mean
RT excerpts of text	Experts	41	**165.8478**	**29.25326**	4.56859
	Laypersons	57	**140.1565**	**40.66134**	5.38573

Table 15 – Mean response time of Experts and Laypeople during trials with excerpts of text

		F	Sig.	t	df	Sig. (2-tailed)	SE	95% Confidence Interval of the Difference	
								Low	Upper
Accuracy artificial fingerprints	Equal variances assumed	4.772	.031	3.452	96	**.001**	25.6913	7.442	10.917
	Equal variances not assumed			3.638	95.99	.000	25.6913	7.062	11.672

Table 16 – T-test of difference of mean response time between Experts and Laypeople during trials with excerpts of text
($p < .05$)

Hypothesis 3: *Both groups of participants will have poorer performances in blocks where contextual information is included before the stimuli, i.e., they will have lower accuracy scores and take more time to complete the task*.

The design in this study allowed analysis of the interactions between independent variables for each of the dependent variables, i.e. to conduct a *stimuli type* X *context* X *type of participant* analysis using a *2-way mixed ANOVA* for each of the dependent variables. However, due to the fact that dependent variable accuracy was not a scale variable but instead categorical within each condition, assumptions for conducting a *2-way mixed design ANOVA* for this variable were not met. Hence, *paired-sample t-tests* were conducted to analyse if there were significant differences in accuracy within blocks where contextual information was included and the control block (without contextual information), and for the dependent variable response time, a *2-way mixed design ANOVA* was conducted to observe interactions. To organise the results, outcomes are presented below, firstly regarding the accuracy, followed by response time.

Accuracy

H.3.1 Fingerprint examiners and Artificial fingerprints

Starting with the fingerprint examiners' answers, a *paired samples test* analysed the differences between accuracy within blocks in the condition "artificial fingerprints". With exception of the "crime type" block, the analysis demonstrated that there were significant differences such that fingerprint examiners were more accurate in the control block when compared with the other two blocks (criminal record and previous conclusion). This can be observed in table 17 below.

				95% Confidence Interval of the Difference				
		Mean	*SD*	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Control – Crime type	.07317	.5192	-.09069	.23703	.902	40	**.372**
Pair 2	Control – Criminal record	.51220	.5967	.32384	.70055	5.496	40	**.000**
Pair 3	Control – Previous conclusion	.87805	.7140	.65269	1.10341	7.875	40	**.000**

Table 17 – Fingerprint examiners | Artificial Fingerprints | Accuracy within blocks *paired-sample t-tests* ($p < .05$)

H.3.2 Fingerprint examiners and Excerpts of text

As showed in table 18, in the condition "Excerpts of text", fingerprint examiners always performed significantly better in the *control block* when compared with blocks that had contextual information included.

				95% Confidence Interval of the Difference				
		Mean	*SD*	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Control – Publication type	-1.87399	.33055	-1.978	-1.76965	-36.302	40	**.000**
Pair 2	Control – Student Record	-2.81098	.49582	-2.967	-2.654	-36.302	40	**.000**
Pair 3	Control – Previous conclusion	-7.00627	1.25587	-7.402	-6.609	-35.722	40	**.000**

Table 18 – Fingerprint examiners | Excerpts of text | Accuracy within blocks *paired-sample t-tests* ($p < .05$)

H.3.3 Laypeople and Artificial fingerprints

Observing the results from laypeople's answers, table 19 demonstrated that the accuracy of these participants in the condition "Artificial fingerprints" was significantly different for two blocks with contextual information (the *criminal record block* and the *previous conclusion block*). Results showed no significant differences between the *control block* and the *crime type block* similar to fingerprint examiners (see H.3.1).

				95% Confidence Interval of the Difference				
		Mean	*SD*	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Control – Crime type	.05263	.789	-.156	.261	.504	56	**.616**
Pair 2	Control – Criminal record	.50877	.984	.247	.769	3.903	56	**.000**
Pair 3	Control – Previous conclusion	.89474	.958	.640	1.148	7.053	56	**.000**

Table 19 – Laypeople | Artificial fingerprints | Accuracy within blocks *paired-sample t-tests* ($p < .05$)

H.3.4 Laypeople and Excerpts of text

In the condition "*Excerpts of text*", laypeople responded significantly better in the *control block* when compared with the *publication type block* and the *previous conclusion block* as opposed to the comparison of their accuracy between the *control block* and the *publication type block* (table 20).

				95% Confidence Interval of the Difference				
		Mean	*SD*	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Control - Publication type	.10526	.880	-.128	.339	.903	56	**.370**
Pair 2	Control – Student record	.59649	.821	.379	.814	5.487	56	**.000**
Pair 3	Control – Previous conclusion	.70175	.844	.478	.926	6.274	56	**.000**

Table 20 – Laypeople | Excerpts of text | Accuracy within blocks *paired-sample t-tests* ($p < .05$)

Response time

H.3.5 Multivariate analysis

In this section, a mixed design ANOVA (table 21) was used to explore the within- and between-subjects influence of the independent variables 'stimuli type (fingerprint/text)', context (3 types, plus control), and expertise (expert/layperson) on the dependent variable 'response time'. This approach enables an interpretation of both the main effects of individual independent variables, as well as any interactions between variables. The factorial design of the experiment resulted in 8 conditions in which to explore the within-subject effects as well as the between-subjects effect of 'expertise'. This type of analysis was only possible for the dependent variable 'response time' (measured as an interval variable), but not for the other dependent variable of interest in this experiment since 'accuracy' scores are categorical within each condition. Results reported here used the Greenhouse-Geisser ε correction values since conditions of sphericity were violated for Mauchly's Test of Sphericity (Field, 2009).

Regarding context, there was a significant main effect of contextual information on response time (F (1.04, 99.96) = 1894.25, p = .001), meaning that response time was significantly slower for all types of context compared to the control (no context) condition.

There was also a main significant effect of type of stimuli on response time (F (1, 96) = 1107.37, p = .001), meaning that response times were slower for the excerpts of text conditions compared to the fingerprint stimuli conditions.

In addition to the main effects of context and stimuli type, three significant interactions were observed. A significant interaction between the type of context and the type of participant was observed (F (1.04, 99.96) = 134.25, p = .001), highlighting that participants differed in their response time according to their type of expertise, and a significant interaction between the type of stimuli and the type of participant was also observed (F (1, 96) = 26.65, p = .001), stressing that response time for both groups of participants differed between the two types of stimuli. Finally, a significant interaction between the type of context X the type of stimuli X the experience of the participant was found (F (1.12, 107.17) = 1512.67, p = .001).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	ηp2
Context	Sphericity Assumed	2448.780	3	816.3	1894.254	**.000**	0.952
	Greenhouse-Geisser	2448.780	1.041	2351.7	1894.254	**.000**	0.952
Context * Type	Sphericity Assumed	173.554	3	57.9	134.253	**.000**	0.583
	Greenhouse-Geisser	173.554	1.041	166.7	134.253	**.000**	0.583
Error (Context)	Sphericity Assumed	124.103	288	0.4			
	Greenhouse-Geisser	124.103	99.963	1.2			
Stimuli	Sphericity Assumed	86.753	1	86.8	1107.365	**.000**	0.920
	Greenhouse-Geisser	86.753	1.000	86.8	1107.365	**.000**	0.920
Stimuli * Type	Sphericity Assumed	2.088	1	2.1	26.646	**.000**	0.217
	Greenhouse-Geisser	2.088	1.000	2.1	26.646	**.000**	0.217
Error (Stimuli)	Sphericity Assumed	7.521	96	0.1			
	Greenhouse-Geisser	7.521	96.000	0.1			
Context * Stimuli	Sphericity Assumed	32.818	3	10.9	1510.696	**.000**	0.940
	Greenhouse-Geisser	32.818	1.116	29.4	1510.696	**.000**	0.940
Context * Stimuli * Type	Sphericity Assumed	32.861	3	10.9	1512.665	**.000**	0.940
	Greenhouse-Geisser	32.861	1.116	29.4	1512.665	**.000**	0.940
Error (Context *Stimuli)	Sphericity Assumed	2.085	288	0.007			
	Greenhouse-Geisser	2.085	107.174	0.019			

Table 21 – Within subjects effects of mixed design ANOVA

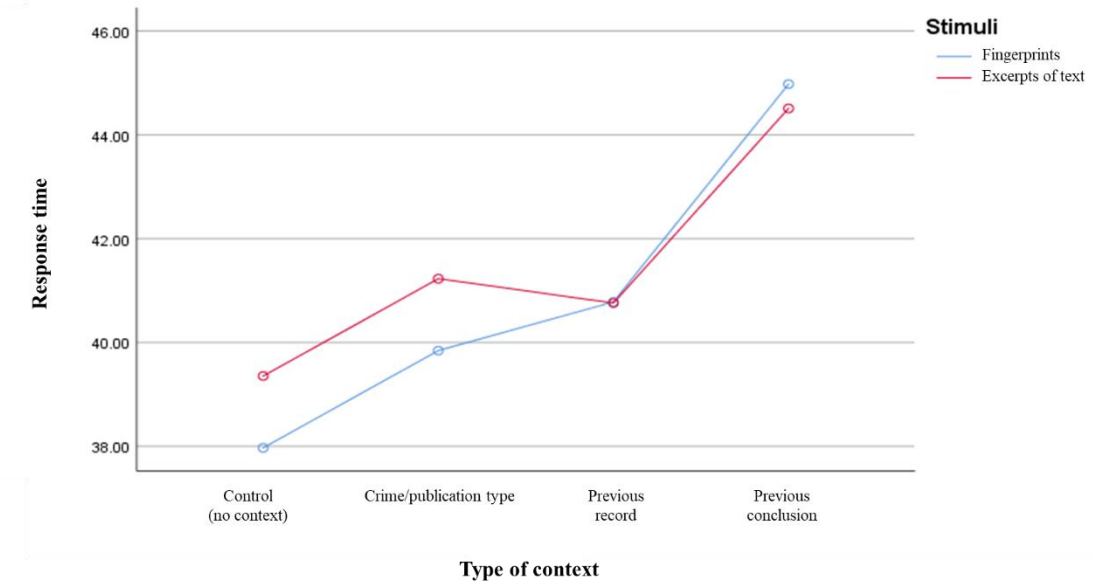

Figure 18 – response time of experts (stimuli type X Context)

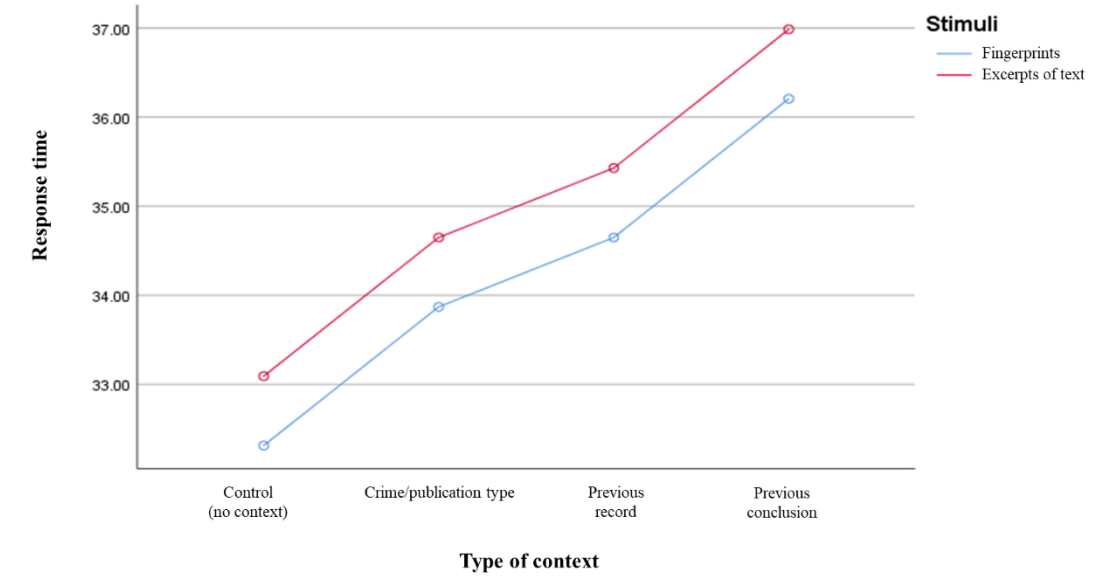

Figure 19 – response time of laypeople (stimuli type X Context)

Regarding the results between experts and laypeople (table 22), the analysis also showed a significant between-subjects effect F (1, 96) = 12.56, p = .001, confirming that response time for experts was slower overall compared to laypeople.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	137112.591	1	137112.591	1694.906	0.000	0.946
Type	1016.056	1	1016.056	12.560	0.001	0.116
Error	7766.101	96	80.897			

Table 22 – Between subjects effects of mixed design ANOVA

3.5 Discussion

In the field of forensics, namely within fingerprint analysis, research has demonstrated that experts exceed in performance when compared with laypeople (Thompson, Tangen & McCarthy, 2013; Busey & Vanderkolk, 2005). Results presented in this chapter were consistent with this rationale, as they demonstrated that fingerprint examiners showed greater accuracy and took more time than laypeople to make their decisions when comparing artificial fingerprints. Taking more time during the task can be considered as one of the characteristics of experts, i.e. being more conservative (Thompson, Tangen & McCarthy, 2013) and investing more time in order to attend to details within stimuli that laypeople do not pay attention to (Shanteau, 1988).

Still related to the performance between experts and laypeople in this study, it is of interest to point out that during the pattern recognition task using excerpts of text, experts were significantly slower than laypeople, however, accuracy between laypeople and experts showed no significant differences. This can suggest what was expected, that this stimulus of excerpts of text was familiar for both groups, and therefore, accuracy was expected to be similar between both groups. Nevertheless, due to the fact that laypeople do not have experience in pattern recognition tasks such as fingerprint analysis in their day-to-day work, results showed a trend that suggested fingerprint examiners to be more accurate (in their mean accuracy), demonstrating abilities that were developed throughout their work as suggested by Weiss & Shanteau (2014).

In the field of fingerprint analysis, Thompson, Tangen and McCarthy (2011) claimed that laypeople can perform at a high level as these individuals were able to correctly identify more than half of the matching pairs in their study. Even though laypeople and forensic practitioners were compared regarding their performance,

this study explored to what extent both groups would have their decisions influenced by typical contextual information that one can easily find during casework within a fingerprint bureau. For both types of stimuli, results retrieved from H3 demonstrated that both groups do suffer from bias due to some types of contextual information, making participants less accurate as well as slower than when they do not have any type of context surrounding the stimuli.

Regarding contextual information, results in this study are partially in accordance with claims related to the potential negative effects that contextual information has within forensic evidence decision-making (Alexander, 2015a; Alexander, 2015b; Osborne, Wood, Kieser and Zajac, 2014; Kassin, Dror & Kukucka, 2013; Dror & Hampikian, 2011; Tangen, Thompson & McCarthy, 2011; Garret & Neufeld 2009) as one of the types of context (i.e. type of the crime/type of publication) showed no significant influence in individuals' accuracy in both groups of participants. There were several significant main effects describing the context to be a variable that had an effect on response time. Hence, even though contextual information seemed to not affect accuracy, it does affect performance partially as it makes decision makers slower when making their decisions.

Results in this study challenge the discussion associated with the influences that contextual information have in individuals' performance. As it was observed, either experts or laypeople had decision making influenced by the presence of contextual information for both types of stimuli. Except for the condition "type of crime/publication", the other two conditions where contextual information was included ("criminal record/student record" and "previous conclusion"), individuals were significantly less accurate and faster when compared with the control condition (no context). This assumption is a first step to suggest further research that will focus on the differences that different contextual information has in individuals' performance as opposed to assuming that all contextual information will negatively influence accuracy. Regarding the effects of this type of contextual information, Chapters 3 and 4 focused on trying to elaborate another perspective to observe it and to discuss its use within fingerprint analysis practices.

3.6 Conclusion

The current chapter observed differences of accuracy and response time between fingerprint examiners and laypeople related to the presence of different contextual information during a pattern recognition task. In this study, two types of stimuli were used; excerpts of text and artificial fingerprints. Excerpts of text enabled laypeople to complete trials using stimuli that were familiar to them as opposed to artificial fingerprints that were familiar only to fingerprint examiners. Results demonstrated that both groups of participants suffered equally from contextual information and that a specific type of contextual information did not influence negatively participants' accuracy, in this case, the type of crime/type of publication. This outcome suggests that some contextual information may have no influence on accuracy, and therefore further studies that focus on this aspect should be encouraged.

4. EFFECTS OF CONTEXTUAL INFORMATION ON PERFORMANCE, NEED FOR COGNITION AND MOTIVATION OF FINGERPRINT EXAMINERS DURING THE VERIFICATION PHASE OF THE ACE-V METHODOLOGY: A TRANSNATIONAL STUDY

4.1 Abstract

The guidelines for fingerprint analysis have been changing regularly since high profile cases of erroneous decisions such as the Brandon Mayfield Case and the McKie Case, where the influences of contextual information have been found to affect the accuracy of decisions. However, studies that have found the negative effects that contextual information has on examiners' performance did not acknowledge the differences between different types of contextual information. In this study, 67 fingerprint examiners, from 15 forensic bureaus, based in 9 countries participated in an experiment that aimed to differentiate the effects of different types of context on their performance. Results suggest that not all of the contextual information within a case has a negative influence on performance. In addition, the type of sample enabled observations regarding methodological approaches and quality standards that were in practice within a range of forensic bureaus.

4.2 Introduction

Forensic science remains dependent on human decision making as a primary tool (Dror & Cole, 2010), which makes it vulnerable to errors of various types. In 2018 in the U.S., the National Registry of Exonerations reported 151 cases where inmates were exonerated due to misconduct of forensic evidence (NRE, n.d.). These cases represent a total of +1639 years that inmates spent in prison erroneously. In those cases, most of the decisions that were conducted within forensic practices were observed to be affected by external sources that influenced them negatively (Kassin et al., 2013).

Even though forensic examiners can be acknowledged as experts due to their experience and knowledge in specific fields (Weiss & Shanteau, 2014), they can

also be overconfident and commit errors even without noticing (Shanteau & Gaeth, 1983; Golde, 1970). Some of the errors that were found related to what Aronson (1999) called the use of 'cognitive misers', i.e., shortcuts that people tend to use to make decisions with minimal effort. This type of phenomenon was first defined by cognitive scientists when describing the effects of cognitive biases (Tversky & Kahnemann, 1974) that have also been observed in forensics in general and specifically within the use of DNA (Dror & Hampikian, 2011; Ask, Rebelius & Granhag, 2008), bitemarks (Osborne, Wood, Kieser & Zajac, 2014), footwear, tool mark and firearms (FSR, 2020), hair and fibres (FSR, 2020) fingerprints (Kassin, Dror & Kukucka, 2013; Langenburg, Champod & Wertheim, 2009), amongst others that were scrutinised in official reports that focused their attention on the human factors within forensics (Executive Office of the President, 2016).

Aiming to minimise potential errors, official guidelines have been released by different agencies such as the National Academy of Science's Report "Strengthening Forensic Science in the United States: A Path Forward" (NAS, 2009), the Report to the President – PCAST Report (Executive Office of the President, 2016), and the Forensic Science Regulator Series (FSR, 2013). Since their release, policymakers have attempted to overcome the issues behind the potential sources of errors, such as the effects of cognitive biases on the performance of forensic practitioners. As a result, the numerical approach of counting minutiae within fingerprint analysis has been questioned in the past (Evett & Williams, 1996), and suggestions such as using likelihood ratios have been presented (Champod, 2009). However, forensic bureaus continue to work with different standards (Gonçalves, 2017) with some fingerprint bureaus operating within a holistic approach while others have maintained a numerical approach.

Related to the effects of cognitive bias on fingerprint analysis, a study conducted by Langenburg, Champod and Wertheim (2009) explored the effects of contextual information specifically during the Verification phase of the ACE-V methodology. The study demonstrated that fingerprint examiners were more conservative when they were exposed to irrelevant information. Although this study was a landmark for the comprehension of the effects of cognitive bias originating from contextual information, some limitations were left to overcome, which this study aims to address.

The sample that participated in the experiment carried by Langenburg, Champod and Wertheim was mainly represented by American fingerprint examiners. In the study described in this chapter, fingerprint examiners from a range of countries participated. Another difference was the fact that the present study did not allow participants to reach an inconclusive decision. This is similar to what Thompson, Tangen and McCarthy did with their forced-decision setup (2011) and to challenge the idea of Biedermann, Bozza, Taroni and Vuille (2019) regarding the value of having inconclusive decisions within forensic science disciplines. The study in this chapter also observed novel variables such as the need for cognition and the motivation that fingerprint examiners have during their work.

4.3 Method

4.3.1 Sample

In this study, the sample consisted of 67 fingerprint examiners, 34 females and 33 males (M_{age} = 39.58 years, SD = 10.41, min = 23, max = 62). Individuals were invited through gatekeepers at 15 forensic bureaus where criminal investigations were carried out regularly, based in 9 countries (UK, Portugal, Belgium, Netherlands, Germany, Brazil, U.S., China, Australia). The only requirement for a fingerprint examiner to participate in this experiment was that the ACE-V process was used during case-work at their bureau within fingerprint identification processes. Forensic bureaus represented in the sample adopted one of two possible approaches to conducting fingerprint identification processes, a holistic approach or a numerical approach. The holistic approach observes the evidence within a holistic manner, whereas the numerical approach requires the fingermark collected at a crime scene and the candidate print that is compared with, to match on a certain number of minutiae points (which differ between bureaus).

In this experiment, at the time of the data collection, 8 of the bureaus followed a holistic approach and 7 bureaus followed a numerical approach. Within the sample, 8 bureaus were accredited and 7 did not have any kind of accreditation. Specific descriptions related to accreditation were kept confidential. A detailed description of these demographics is provided below in table 23.

Country	Bureau	No. participants	Approach	Accredited
UK	Bureau 1	7	Holistic	No
	Bureau 2	3		No
	Bureau 3	7		Yes
Portugal	Bureau 4	12	Numerical	No
Belgium	Bureau 5	9	Numerical	No
Netherlands	Bureau 6	2	Holistic	Yes
Germany	Bureau 7	2	Numerical	Yes
Brazil	Bureau 8	1	Numerical	No
	Bureau 9	1		
	Bureau 10	1		
U.S.	Bureau 11	1	Holistic	Yes
	Bureau 12	2		
	Bureau 13	4		
China	Bureau 14	9	Numerical	Yes
Australia	Bureau 15	6	Holistic	Yes

Table 23 – demographic information related to bureaus where participants worked

Fingerprint examiners that performed the task in this study had on average 10.61 years of experience (s = 10.61, min = 1, max = 35) at the time they participated in the experiment. Since training is not standardised globally, there were different types of criteria for an individual to be recognised as an independent examiner depending on the bureau where the examiner worked. Due to that, consultation with 5 IAI certified senior fingerprint examiners and academics who work within the field of fingerprints was conducted to develop a scale in order to organise participants by their experience. The scale was suggested to split individuals into 3 categories. The number of examiners in each category is shown in table 24.

Rank	Trainee	Examiner	Senior Examiners
Years of experience	Less than 2	Between 2 and 7	More than 7
Number of participants	5	22	40

Table 24 – Ranks of experience of individuals

4.3.2 Design

This study was conducted on-site in forensic bureaus using the researcher's laptop via the software OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) and Python programming language. Data was subsequently analysed using SPSS 25.0 (IBM, 2017).

Participants were presented with the purpose of the study during a first meeting arranged by gatekeepers who generally were bureau managers. Initially, participants were told that the study aimed to explore the effects of different types of cognitive biases on their work. It was also explained that the study would be carried out on-site at the fingerprint examiners' workplace running on the researcher's laptop in order to ensure consistency confidentiality and anonymity of results.

Before participation, individuals were told about the ethical implications of the study. It was explained that during the experiment there were no harms expected to participants, however, that they could stop the experiment at any point without any kind of personal consequences. Ethical approval was obtained by the University of Leicester (Appendix D) and at some of the forensic bureaus that required additional ethical clearance. Also, previous to entering the premises, at some of the forensic bureaus, a personal check of the researcher was conducted by the bureau due to security measures. Prior to participation, individuals signed the informed consent for their participation, where it was also mentioned that no individual results would be shared with any person from their bureau or be accessible by anyone other than the researcher and the supervisors of this thesis.

Within this study, a pattern recognition task was designed to mimic a fingerprint comparison task during ACE-V, specifically during the verification phase. Artificial fingerprints were created and used as stimuli as described in the previous chapter (Chapter 3). Throughout the experiment, individuals were presented with pairs of fingerprints and asked to decide whether the pairs matched or did not match. Each trial presented a simulation of a latent print, mimicking a crime scene mark, and a potential candidate print, simulating a fingermark retrieved from an AFIS system or a ten-print card of a suspect under custody.

Participants had a maximum of 30 seconds to observe the pair of stimuli and make their decision. A pool of five IAI certified senior fingerprint examiners (> 7 years of experience) based in 3 countries were provided with 5 pairs of stimuli similar to those used in this study. They were asked how much time they needed to compare each pair. On average, they responded that a first assessment took them 27 seconds to finish.

Similar to Chapter 3, in this study, comparisons were made within a forced decision-making structure that was used in previous literature (Thompson, Tangen & McCarthy, 2013) where participants needed to decide either by a match or a non-match in their answer, i.e. they could not leave their answer in blank or decide an inconclusive answer.

A set of four blocks was presented through a computer-based experiment. Each block was presented with a specific type of contextual information except for one block that did not have any contextual information and was used as a control block. Within each block where contextual information was included, there were two categories of contextual information, each category had equal weight within the block (i.e. 50% trials within each category). Types of contextual information are described in table 25.

Type of context	Type of manipulation
Type of crime	Volume crime (ex. Burglary)
	Major crime (rape and murder)
Criminal Record	Suspect with a criminal record
	Suspect without a criminal record
Previous conclusion from another FP expert	Correct previous conclusion
	Incorrect previous conclusion
No context	Not applicable / Control block

Table 25 – Manipulation of contextual information in each block of trials

In order to prevent sequential bias, all blocks and trials within each block were randomly assigned (Viera & Bangdiwala, 2007). After a slide with the experiment's guidelines, a warm-up phase was deployed where participants needed to complete 5 comparisons in order to be familiar with the stimuli and the setup. There was no contextual information within the warm-up phase or time constraints.

Following the warm-up phase, indications that the experiment was going to start were presented. Participants were told that they would have trials with and without contextual information, however, they did not know what kind of contextual information would be used. They were also told that there was no time limit to see the contextual information, but there was a time limit of 30 seconds to make their decision. After the first two blocks, a break of 10 minutes was deployed. Two final

blocks were completed followed by a final slide and some time for any questions that participants might have.

At the end of the experiment, two psychometric scales were completed: The Need for Cognition Scale (Cacioppo, Petty & Kao, 1984) and Work Extrinsic and Intrinsic Motivational Scale (Tremblay, Blanchard, Taylor, Pelletier & Villeneuve, 2009).

Due to the fact that not all participants were fluent in English, all of the instructions and material throughout the experiment, as well as the questionnaires that were completed at the end of the experiment were translated into three other languages (French, Chinese, Portuguese). Each translation was conducted by asking two native speakers of the translated language who were also proficient in English to translate the contents of the experiment. As suggested by Sperber (2004), after the first translation, a back-translation was carried out and then compared with the original one in order to observe any significant differences.

A total of 5,628 trials were compared within this study. In each block, there were 21 trials composed of pairs of artificial fingerprints. Proportions of the different patterns of the stimuli (loops, whorls and arches) were adopted from the percentages in the real world and assigned to each block as mentioned by Smith and Bond (2015) – 65% loops, 35% whorls, and 5% arches.

Hypotheses and research questions in this study tested the relationships between four dependent variables: (1) accuracy, (2) response time, (3) need for cognition score and (4) motivation score, and six independent variables that were manipulated: (1) presence of contextual information, (2) years of experience, (3) patterns of artificial fingerprints, (4) type of comparison (i.e. match, non-match and close non-match), (5) type of methodological approach within the bureau (numerical or holistic), (6) accreditation.

Hypotheses

1. Accuracy

- H1.1: *Fingerprint examiners will be less accurate in trials where contextual information is included*
- H1.2.: *Senior fingerprint examiners will be more accurate overall than examiners and trainees*
- H1.3: *Fingerprint examiners that follow a holistic approach within their bureau during their casework will be more accurate than fingerprint examiners that follow a numerical approach*
- H1.4: *Fingerprint examiners who work within an accredited bureau will be more accurate than fingerprint examiners that work within a bureau without accreditation*

2. Response time

- H2.1: *Fingerprint examiners will be slower in trials where contextual information is included*
- H2.2: *Senior fingerprint examiners will be faster than examiners and trainees*
- H2.3: *Fingerprint examiners that follow a holistic approach within their bureau during their casework will be faster than fingerprint examiners that follow a numerical approach*
- H2.4: *Fingerprint examiners who work within an accredited bureau will be faster than fingerprint examiners that work within a bureau without accreditation*

Exploratory questions

A. Accuracy

- EQ A.1.: *Do fingerprint examiners have the same accuracy when comparing non-matches than other types of comparisons such as matches and close non-matches?*

B. Response time

- EQ B.1.: *Do fingerprint examiners spend the same time comparing non-matches than other types of comparisons such as matches and close non-matches?*

C. Need for Cognition

- EQ C.1.: *Will senior fingerprint examiners have higher levels of need for cognition than examiners and trainees?*
- EQ C.2.: *Will fingerprint examiners that follow a holistic approach have higher levels of need for cognition than fingerprint examiners that work within a numerical approach?*
- EQ C.3.: *Will fingerprint examiners that work within an accredited bureau have higher levels of need for cognition?*

D. Motivation

- EQ D.1.: *Are senior fingerprint examiners intrinsically more motivated than examiners and trainees?*
- EQ D.2.: *Will fingerprint examiners that follow a holistic approach have higher levels of intrinsic motivation than fingerprint examiners that work within a numerical approach?*
- EQ D.3.: *Are fingerprint examiners that work within an accredited bureau intrinsically more motivated?*

4.3.3 Materials

4.3.3.1 Stimuli

Similar to study 1, a set of artificial fingerprints were created using SFinGe (Cappelli, 2015). The set of artificial fingerprints consisted of simulations of fingerprints to mimic fingermarks that are found at crime scenes and simulations of scanned potential candidates that can be retrieved from an electronic system such as an AFIS system or ten-print cards of suspects under custody.

In total, 168 artificial fingerprints were created (84 pairs). In order to ensure ecological validity (Shadish, Cook & Campbell, 2002) of the stimuli, the set was sent to five IAI certified senior fingerprint examiners (all > 7 years of experience). All the senior fingerprint examiners agreed that the set was realistic enough to be used within this study.

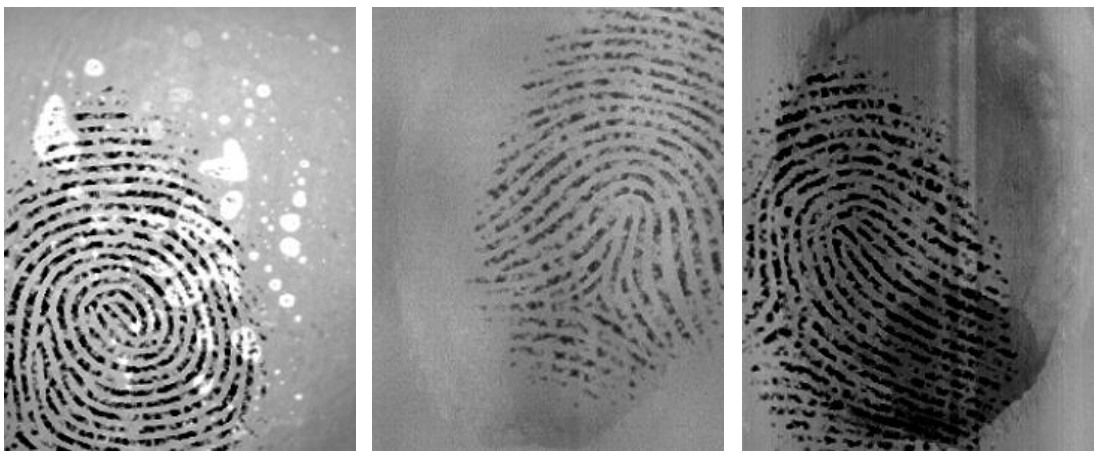

Figure 20. Examples of three artificial generated fingerprint patterns (whorl, arch, loop)

Artificial fingerprints were created in pairs with three possible outcomes: match, non-match and close non-match. Close non-match pairs were named after

discussing the concept of *highly similar non-matches* with fingerprint examiners and with researchers in the field. Close non-match pairs were considered different from standard non-matching pairs as these can mimic highly difficult comparisons such as the case of Brandon Mayfield (Kershaw & Lichtblau, 2004) or the Shirley McKie case (BBC, 2011). During conversations with fingerprint examiners and observation of real casework within forensic bureaus, it was possible to understand that close non-matches were less frequent and therefore the proportions regarding the type of comparison are summarised in table 26 and were suggested and validated by the pool of fingerprint examiners that advised the research in this thesis.

Type of comparison	Estimates in the real world
Match	40%
Non-match	40%
Close non-match	20%

Table 26 – Estimate of different comparisons conducted in the real world

As described in study 1, all pairs of fingerprints were artificially created. This enabled a ground truth dataset where correct answers were known instead of having a ground truth dataset by proxy. This type of setup has also been used in the field of fingerprint comparisons (Mikaelyan & Bigun, 2012) as well as in other fields Roussev (2011).

Figure 21. Example of a matching pair

Figure 22. Example of a non-matching pair

Figure 23. Example of close non-matching pair

Slides containing contextual information were created by combining real information and fictional text. Contextual information was included between trials.

Figure 24. Example of contextual information (Type of crime)

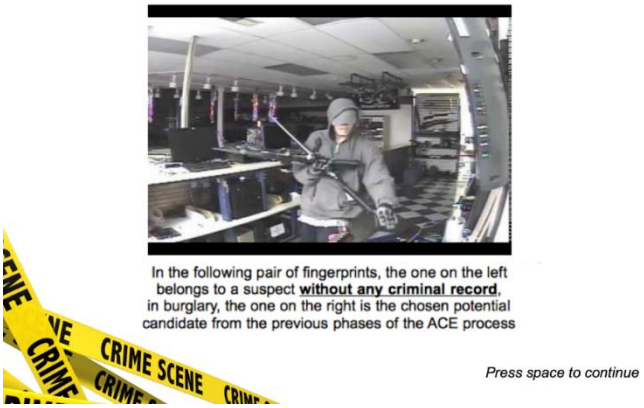

Figure 25. Example of contextual information (suspect's record)

Figure 26. Example of contextual information (Previous Conclusion by another examiner)

4.3.3.2 Need for Cognition Scale

The Need for Cognition Scale (NCS) is a psychometric tool that assesses the level of enjoyment an individual has when engaging in a thinking task such as a decision-making process (Cacioppo, Petty & Kao, 1984). The inventory was designed from the work developed by Cohen, Stotland and Wolfe (1955) who described the term 'Need for Cognition' as a phenomenon occurring when a person completes a task that gives him/her cognitive satisfaction in order to differentiate that kind of satisfaction from other needs. Cacioppo and Petty (1982) identified that no instrument could measure the level of cognition one had when performing a task and created the first inventory, at the time with 34 items.

Even though the questionnaire has not been used to study the level of cognition of experts in the forensic science domain, it has been applied to other populations such as Clinical Psychology (Sadowski & Cogburn, 1997; Bagby, Taylor, & Ryan, 1986), Social Psychology (Wolf, von Hecker, & Maio, 2017; Aquino, Haddock, Maio, Wolf, & Alparone, 2016), Journalism (Liu & Eveland, 2005), Management (Kearney, Gebert, & Voelpel, 2009) and Marketing (Haddock, Maio, Arnold, & Huskinson, 2008). Although researchers have been working towards a smaller version of the scale with only 6 items (Coelho, Hanel & Wolf, 2018), the NCS version of the scale that was applied was the version that consists of 18 items developed by Cacioppo, Petty and Kao (1984), which showed valid stability of its assessment and internal consistency of items (Juric, 2016).

This inventory has previously been applied to individuals from different countries such as Australia (Forsterlee & Ho, 1999), Germany (Bless, Wänke, Bohner, Fellhauer, & Schwarz, 1994), Taiwan (Kao, 1994), Portugal (Silva & Garcia-Marques, 2013), Netherlands (Pieters, Verplanken, & Modde, 1987), Brazil (Gouveia, Mendes, Soares, Monteiro, & Santos, 2015), and in a U.S.-Hispanic sample (Culhane, Morera, & Hosch, 2004), where it showed high internal consistencies, with reliabilities generally varying between $\alpha = .80$ and $\alpha = .90$, as well as stable results across age groups (Soubelet & Salthouse, 2016). Due to that, it was expected that the scale could be applied to individuals from different countries that were part of the sample in this study.

One of the main reasons for the application of this scale within this study was the fact that the NCS has observed a meaningful relationship with a variable that is also the focus of this thesis, motivation. Cacioppo, Petty, Feinstein and Jarvis (1996) found a meaningful relationship with intrinsic motivation of individuals who have high levels of Need for Cognition who showed that they engage in the task due to the cognitive challenge. On the other hand, individuals with low levels of Need for Cognition usually suffer from biased decision-making. Coelho, Hanel and Wolf (2018) also described that individuals who show low levels of Need for Cognition tend to use other sources of decision-making such as heuristics in order to "make sense of the world" (p. 1).

Since this study aimed to analyse how fingerprint examiners are affected by cognitive bias that has its origin from contextual information, the use of such a tool like the NCS was used to recognise and better understand the cognitive flow that examiners had during the tasks they were asked to carry out.

4.3.3.3 Work Extrinsic and Intrinsic Motivation Scale

The Work Extrinsic and Intrinsic Motivation Scale (WEIMS) is an 18-item psychometric instrument that aims to assess the work motivation of individuals by analysing and differentiating their intrinsic and extrinsic motivation (Tremblay, Blanchard, Pelletier and Villeneuve, 2009). The tool was developed under the grounded theory of self-determination (SDT) claimed by Deci and Ryan (2000) and has been used to measure work motivation in different settings such as Occupational Therapy (Chai, Teoh, Razaob & Kadar, 2017), Management (Nordhall & Knez, 2018; Shu, 2015) and Social Work (Proença & Cristina, 2013).

Tremblay et al. (2009) described the instrument to be composed of six subscales that explain the two different types of motivation, namely extrinsic motivation (EM) and intrinsic motivation (IM). Extrinsic motivation, referring to the act that leads individuals to do something in order to achieve a separable outcome (e.g. reward) (Gagné & Deci, 2005) is composed of four subscales of the WEIMS. The first subscale, Integrated Regulation (INTEG), refers to the act of identifying with the value of an activity to the point that it becomes part of one's sense of self. Secondly, there is the Identified Regulation (IDEN) subscale, which is related to the action of doing an activity as if it was one's own due to the identification one has with its meaning and value. The third subscale, Introjected Regulation (INTRO), observes the regulation of behaviour through self-worth contingencies such as self-esteem and the fourth subscale, External Regulation (ER), explains the motivation that one has related to an external outcome that is expected. On the other side of the spectrum, there is Intrinsic Motivation (IM) which is assessed by its subscale within WEIMS. There is also a subscale designated by Amotivation (AMO), which illustrates how individuals lack the intention to act or act passively. This Amotivation subscale is accounted in contrast to either the assessment of EM or IM since both types of motivation are intentional (Chai et al., 2017).

In order to describe the levels of EM and IM, Tremblay et al. (2009) suggested using the mean score of each of the subscales that compose each type of motivation. For EM, one should analyse the sum score of the means of the four subscales that structure that type of motivation, i.e. EM = $\Sigma\ [x_{INTEG}+x_{IDEN}+x_{INTRO}+x_{EXT}]$. IM is obtained by the mean score of subscale IM.

Intrinsically motivated people are described as having enjoyment for their work only due to the specifications of the tasks. On the other hand, extrinsically motivated people have their motivations influenced by more than only the tasks they carry out such as external sources like rewards (Shu, 2015).

An index that is possible to retrieve from WEIMS is the Work Self-Determined Index (W-SDI). The W-SDI has been used to identify whether individuals present a self-determined or a nonself-determined motivational profile (Chai et al., 2017). The W-SDI is found by multiplying the mean of each subscale by weights corresponding to the underlying level of self-determination (Chai et al., 2017). The formula for determining the W–SDI is $[W\text{–}SDI = (3 \times x_{IM}) + (2 \times x_{INTEG}) + (1 \times x_{IDEN}) + (-1 \times x_{INTRO}) + (-2 \times x_{EXT}) + (-3 \times x_{AMO})]$ (Chai et al., 2017).

Positive scores of W-SDI indicate that an individual is primarily internally driven, whereas negative scores are an indication that the individuals have their work motivation shaped by extrinsic factors (Shu, 2015). Nordhall and Knez (2018) claimed that the higher the levels of intrinsic motivation individuals had, the greater they would relate to their work, and therefore they were expected to have better performance.

Hoffman (2015) also described the two types of behaviour that individuals can demonstrate regarding their type of motivation that is mostly present when carrying out a task. According to the author, intrinsically motivated individuals (identified as Mastery oriented individuals), have their personal goals maintained by intrinsic self-improvement. On the other hand, extrinsically motivated individuals (defined as Normative oriented), seek motivation from other sources than the stimuli of their task, thus they can be more influenced by external factors and commit erroneous decisions more often when carrying specific tasks.

The results retrieved from the WEIMS allowed a better understanding of two main points. Firstly, it was possible to observe if fingerprint examiners who were mostly intrinsically motivated also had higher levels of Need for Cognition, and secondly, this instrument allowed the comprehension of the type of motivation that was more susceptible to be biased during tasks similar to the casework that fingerprint examiners experience.

4.4 Results

In this section, hypotheses have been tested individually to facilitate the comprehension of the statistical work conducted. An overall discussion follows this section of results.

4.4.1. Accuracy

Kolmogorov-Smirnov normality test for variable 'Accuracy' within the different blocks was conducted. It was not possible to assume a normal distribution within this variable in general or in any of the blocks ($p < .05$). Thus, non-parametric tests have been conducted in order to test the hypotheses related to this variable.

H1.1: Fingerprint examiners will be less accurate in all trials where contextual information is included

Table 27 showed that participants had higher median scores ($\chi = 10.10$, $s = 2.35$) when there was no contextual information presented before the trials with information related to the type of the crime ($\chi = 9.61$, $s = 2.73$), the criminal record ($\chi = 8.64$, $s = 2.34$) or the previous conclusions ($\chi = 8.51$, $s = 2.01$). Since the variable 'Accuracy' was not normally distributed, a Wilcoxon Signed-rank test was conducted to observe whether there were differences in accuracy within the four blocks.

	n	x	s	Min	Max	Percentiles		
						25th	50th (Median)	75th
Accuracy "no context/control	67	**10.10**	2.349	6.00	17.00	8.0000	10.0000	11.00
Accuracy "type of crime"	67	**9.612**	2.730	6.00	18.00	8.0000	9.0000	11.00
Accuracy "criminal record"	67	**8.642**	2.346	3.00	17.00	7.0000	8.0000	10.00
Accuracy "previous conclusion"	67	**8.506**	2.010	4.00	15.00	7.0000	8.0000	9.00

Table 27 – Accuracy mean scores within blocks with different contextual information

Table 28 shows that when participants were performing the experiment without being presented with contextual information they showed higher accuracy when compared with the block where the type of the crime was provided (37 subjects performed better), the block where the criminal record was given (43 subjects showed more correct answers) and the block where a previous conclusion was showed (47 individuals had greater accuracy).

		n	Mean Rank	Sum of Ranks
"type of crime" VS "no context / control"	Negative Ranks	**37**	28.18	1042.50
	Positive Ranks	20	30.53	610.50
	Ties	10		
"criminal record" VS "no context / control"	Negative Ranks	**43**	30.24	1300.50
	Positive Ranks	12	19.96	239.50
	Ties	12		
"previous conclusion" VS "no context / control"	Negative Ranks	**47**	32.85	1544.00
	Positive Ranks	14	24.79	347.00
	Ties	6		

Table 28 – Ranks of accuracy within different types of contextual information

Except for the comparison between the block without contextual information and the block with the type of the crime (z = -1.73, *p* = .083), comparisons were statistically significant. As described (table 29), a statistically significant difference between medians in the block without contextual information and the blocks with the criminal record (z = -4.48, p < .001) and the previous conclusions (z = -4.33, p < .001) were observed. Therefore, H1 was partially rejected as there was no statistically significant difference between the trials within the control block and the block where information related to the type of the crime was included.

	"type of crime" VS "no context/control"	"criminal record" VS "no context/control"	"previous conclusion" VS "no context/control"
Z	-1.733[b]	-4.483[b]	-4.335[b]
Asymp. Sig. (2-tailed)	**.83**	.001	.001

Table 29 - Wilcoxon's signed-rank test "*accuracy within blocks*"
(*p* < .05)

H1.2.: *Senior fingerprint examiners will be more accurate overall than examiners and trainees*

Hypothesis H1.2. analysed the differences in accuracy based on the type of experience that participants had. Since the variable accuracy did not have a normal distribution a non-parametric test was conducted, in this case, the Kruskal-Wallis H Test. As table 30 described, when observing accuracy of trainees (*n* = 5), examiners (*n* = 22) and senior examiners (*n* = 40), scores were not significantly different ($\chi^2(2) = 1.311$, *p* = .519), therefore H1.2. was rejected.

Kruskal-Wallis H	1.311
df	2
Asymp. Sig.	**.519**

Table 30 – Kruskal-Wallis H Test for accuracy between ranks of experience
(*p* < .05)

H1.3: *Fingerprint examiners that follow a holistic approach within their bureau during their casework will be more accurate than fingerprint examiners that follow a numerical approach*

To differentiate accuracy of different work approaches that fingerprint examiners follow, a Kruskal-Wallis test was performed to compare the accuracy of individuals who work within a holistic approach (n = 32) and from a numerical approach (n = 35). Even though trends of means (table 31) showed that fingerprint examiners who work within a numerical approach had higher accuracy ($\bar{x}$ = 36.67) than individuals who worked within a holistic approach ($\bar{x}$ = 31.08), differences were not significant ($\chi^2(1) = 1.384$, $p = .239$) (table 32), therefore H1.4 was rejected.

	n	Mean Rank
Numerical	35	**36.67**
Holistic	32	**31.08**
Total	67	

Table 31 – Mean accuracy between methodological approaches

Kruskal-Wallis H	1.384
df	1
Asymp. Sig.	.239

Table 32 – Kruskal-Wallis H Test for accuracy between methodological approach ($p < .05$)

H1.4: *Fingerprint examiners who work within an accredited bureau will be more accurate than fingerprint examiners that work within a bureau without accreditation*

Aiming to observe differences that accreditation could have in individuals' performance within fingerprints bureaus, a Kruskal-Wallis test was conducted to analyse differences of participants' accuracy who worked at accredited bureaus (n = 33) from those who worked at non-accredited bureaus (n = 34). Participants from accredited bureaus did show a lower trend in their accuracy mean ($\bar{x}$ = 32.45) than

fingerprint examiners who worked at non-accredited bureaus (χ = 35.50) (table 33), however, the test showed no statistically significant differences between both groups ($\chi^2(1) = .411$, $p = .521$) (table 34).

	n	Mean Rank
Not accredited	33	32.45
Accredited	34	35.50
Total	67	

Table 33 – Mean accuracy scores between different levels of experience

Kruskal-Wallis H	.411
df	1
Asymp. Sig.	.521

Table 34 – Kruskal-Wallis H Test for accredited and non-accredited bureaus ($p < .05$)

4.4.2. Response time

In order to analyse if the variable 'Response time' had a normal distribution within the four blocks, a Kolmogorov-Smirnov test of normality was conducted. Except for the control block (no contextual information), the variable showed no normality within its distribution ($p < .05$), therefore, non-parametric tests for hypotheses within H2 were used.

H2.1.: *Fingerprint examiners will be slower in trials where contextual information is included*

For H2.1. the Wilcoxon-Signed Rank Test (table 36) was performed which elicited significant differences between the block without contextual information and all of the other three blocks, meaning that individuals were significantly faster when responding to trials without any type of contextual information (χ_{tr} = 396.41 seconds, s = 115.73) than when carrying out the tasks knowing the type of the crime (Z = -4.213, p < .05, χ_{tr} = 430.56, s = 116.85), the criminal record of the suspect

(Z = -5.341, p < .05, x_{tr} = 444.08, *s* = 115.89) or when they knew a previous conclusion from other examiner (Z = -5.547, p < .05, x_{tr} = 464.47, *s* = 111.03). H2.1. was therefore accepted.

	n	*X*	*S*	Min	Max	Percentiles		
						25th	50th (Median)	75th
RT "no context/control"	67	**396.41**	115.73	109.30	601.07	301.06	399.76	502.30
RT "type of crime"	67	**430.56**	116.85	138.56	610.45	345.40	453.80	525.07
RT "criminal record"	67	**444.08**	115.89	143.33	626.62	353.58	463.90	530.53
RT "previous conclusion"	67	**464.47**	111.03	163.82	615.33	372.59	488.01	555.25

Table 35 – Mean response times within blocks with different contextual information

	"type of crime" VS "no context/control"	"criminal record" VS "no context/control"	"previous conclusion" VS "no context/control"
Z	**-4.213**	**-5.341**	**-5.547**
Asymp. Sig. (2-tailed)	**.001**	**.001**	**.001**

Table 36 - Wilcoxon's signed-rank test for response times within blocks
($p < .05$)

H2.2: *Senior fingerprint examiners will be faster than examiners and trainees*

H2.2. analysed whether there were significant differences in response times between different levels of experience. A Kruskal-Wallis H test was performed (table 37), demonstrating that there were no significant differences regarding time response distribution scores across the three levels of experience ($x^2(2) = .437$, $p = .804$). Therefore, H2.2. was rejected.

Kruskal-Wallis H	.437
df	2
Asymp. Sig.	.804

Table 37 – Kruskal-Wallis H Test for response times between different levels of experience
($p < .05$)

H2.3: *Fingerprint examiners that follow a holistic approach within their bureau during their casework will be faster than fingerprint examiners that follow a numerical approach*

Regarding the differences in response time that individuals with different methodological approaches have, table 38 shows the results from the Kruskal-Walis H Test that described individuals using a numerical approach ($n = 35$, $x = 27.97$) being significantly faster ($x^2 2(1) = 7.015$, $p = .008$) than the participants using a holistic approach ($n = 32$, $x = 40.59$). Thus, H2.3. was rejected.

	n	Mean Rank
Numerical	35	**27.97**
Holistic	32	**40.59**
Total	67	

Table 38 – Mean response times between methodological approaches

Kruskal-Wallis H	**7.015**
df	1
Asymp. Sig.	**.008**

Table 39 – Kruskal-Wallis H Test for response times between methodological approach
($p < .05$)

H2.4: *Fingerprint examiners who work within an accredited bureau will be faster than fingerprint examiners that work within a bureau without accreditation*

A Kruskal-Walis H Test (table 40) showed that fingerprint examiners who worked within accredited bureaus did not have a significantly different response time than examiners from non-accredited fingerprint bureaus ($X^2(1) = 3.261$, $p = .071$). It was found a mean rank time response score of 29.76 for non-accredited bureaus and 38.36 for accredited bureaus. For that reason, H2.4. was rejected.

	n	Mean Rank
Not accredited	34	**29.76**
Accredited	33	**38.36**
Total	67	

Table 40 – Mean response times between accredited and non-accredited bureaus

Kruskal-Wallis H	**3.261**
df	1
Asymp. Sig.	**.071**

Table 41 – Kruskal-Wallis H Test for response times between accredited and non-accredited bureaus
($p < .05$)

4.4.3. Exploratory Research Questions

A. Accuracy

RQ A.1: *Do fingerprint examiners have the same accuracy when comparing non-matches than other types of comparisons such as matches and close non-matches?*

Non-matching pairs had a mean score of .51 correct answers, whereas matching pairs showed a mean score of .73 correct answers. Close non-matching pairs were the trials where individuals had lowest accuracy ($x = .33$) (table 42).

	n	*X*	*S*	Min	Max	Percentiles		
						25th	50th (Median)	75th
Accuracy "non-match pairs"	67	**.5081**	.12185	.33	.92	.4167	.4583	.5833
Accuracy "match pairs"	67	**.7301**	.14220	.25	.92	.6667	.7500	.8333
Accuracy "close non-match pairs"	67	**.3315**	.12831	.17	.79	.2500	.2917	.3958

Table 42 - Ranks of accuracy within different types of pairs regarding matching or non-matching outcomes

To observe whether there were significant differences in accuracy when fingerprint examiners were analysing non-matching, matching or close non-matching pairs, a Wilcoxon signed-rank test (table 43) showed that there were statistically significant differences between matching pairs and non-matching pairs (Z = -5.605, $p < .005$) and between close non-matching and non-matching (Z = -6.995, $p < .005$). Individuals had higher accuracy for non-matching pairs compared with close non-matching, but lower accuracy for non-matching pairs compared to matching pairs.

	"match pairs" VS "non-match pairs"	"close non-match pairs" VS "non-match"
Z	**-5.605**	**-6.995**
Asymp. Sig. (2-tailed)	.001	.001

Table 43 - Wilcoxon's signed-rank test *accuracy within the type of comparison* ($p < .05$)

B. Response time

RQ B.1: *Do fingerprint examiners spend the same time comparing non-matches than other types of comparisons such as matches and close non-matches?*

A Wilcoxon-Signed Rank Test was conducted to observe if individuals were faster or slower for the different types of comparison. Table 45 showed significant differences only between non-matches (x_{tr} = 20.87 seconds, s = 5.13) and matches (x_{tr} = 20.16 seconds, s = 5.41) (Z = -2.617, p < .05), suggesting that participants took longer to complete non-matching trials than matching trials.

	n	*X*	*S*	Min	Max	Percentiles		
						25th	50th (Median)	75th
RT "non-match pairs"	67	**20.87**	**5.13**	7.68	29.25	16.95	22.03	24.99
RT "match pairs"	67	**20.16**	**5.41**	6.34	28.92	15.29	20.57	24.90
RT "close non-match pairs"	67	20.68	5.18	6.37	28.01	15.97	21.70	24.48

Table 44 – Response times within blocks with different contextual information

	"match pairs" VS "non-match pairs"	"close non-match pairs" VS "non-match pairs"
Z	**-2.617**	-1.299
Asymp. Sig. (2-tailed)	**.009**	.194

Table 45 - Wilcoxon's signed-rank test "*RT type of comparison*"
($p < .05$)

C. Need for Cognition

RQ C.1.: *Will senior fingerprint examiners have higher levels of need for cognition than examiners and trainees?*

This section explores the relationship between Need for Cognition scores and different ranks of experience. A Kruskal-Wallis H test was performed (table 46), which demonstrated that there were no significant differences in the Need for Cognition scores across the three different ranks of experience ($X^2(2) = 1.862$, $p = .394$).

Kruskal-Wallis H	**1.862**
df	2
Asymp. Sig.	**.394**

Table 46 – Kruskal-Wallis H Test for NCS between levels of experience
($p < .05$)

RQ C.2: *Will fingerprint examiners that follow a holistic approach have higher levels of Need for Cognition than fingerprint examiners that work within a numerical approach?*

In order to identify if individuals in either of the methodological approaches presented a higher level of Need for Cognition, a Kruskal-Wallis H test was conducted. The test (table 47) showed that there were no differences between individuals that followed a holistic approach or a numerical approach ($\chi 2(1) = .210$, $p = .647$).

Kruskal-Wallis H	**.210**
df	1
Asymp. Sig.	**.647**

Table 47 – Kruskal-Wallis H Test for NfCS between methodological approaches ($p < .05$)

RQ C.3: *Will fingerprint examiners that work within an accredited bureau have higher levels of need for cognition*

A Kruskal-Wallis H test (table 48) was conducted to observe whether the variable Need for Cognition was significantly different between fingerprint examiners that were working in an accredited bureau or a non-accredited bureau. The test showed that there were no significant differences between individuals regarding the accreditation of their bureau ($\chi 2(1) = .021$, $p = .885$).

Kruskal-Wallis H	**.021**
df	1
Asymp. Sig.	**.885**

Table 48 – Kruskal-Wallis H Test for NCS between accredited and non-accredited bureaus ($p < .05$)

D. Motivation

RQ D.1: *Are senior fingerprint examiners intrinsically more motivated than examiners and trainees?*

Aiming to observe if the intrinsic motivation levels were significantly different between the different ranks of fingerprint examiners, a Kruskal-Wallis H Test (table 49) was conducted, which demonstrated that there were no differences between groups in the three ranks of experience ($X^2(2) = .770$, $p = .681$).

Kruskal-Wallis H	**.770**
df	2
Asymp. Sig.	**.681**

Table 49 – Kruskal-Wallis H Test for intrinsic motivation between levels of experience ($p < .05$)

RQ D.2: *Will fingerprint examiners that follow a holistic approach have higher levels of intrinsic motivation than fingerprint examiners that work within a numerical approach?*

The levels of motivation of fingerprint examiners that follow different methodological approaches were verified within a Kruskal-Wallis H Test (table 50), which showed that there were no significant differences between the two types of approach ($X^2(1) = 1.882$, $p = .170$).

Kruskal-Wallis H	**1.882**
df	1
Asymp. Sig.	**.170**

Table 50 – Kruskal-Wallis H Test for intrinsic motivation between methodological approaches ($p < .05$)

RQ D.3: *Are fingerprint examiners that work within an accredited bureau intrinsically more motivated?*

Aiming to observe if the motivation levels were significantly different between the two types of bureaus regarding their accreditation, a Kruskal-Wallis H Test (table 51) was performed. The test showed no differences between accredited and non-accredited groups ($X^2(1) = .054$, $p = .817$).

Kruskal-Wallis H	**.054**
df	1
Asymp. Sig.	**.817**

Table 51 – Kruskal-Wallis H Test for intrinsic motivation between accredited and non-accredited bureaus
($p < .05$)

4.5 Discussion

Continuing the work conducted in Chapter 3, this study focused only on fingerprint examiners. Participants were asked to carry out a task similar to their case-work where it was possible to identify the effects of different types of contextual information on their accuracy and response times.

Previous research has demonstrated that contextual information can influence performance during fingerprint examination processes (Earwaker, Morgan, Harris & Hall, 2015; Dror, Kassin & Kukucka, 2013; Langenburg, Champod & Wertheim, 2009). Since academic claims in that regard have been accepted by the forensic community, organisations have been publishing guidelines (Executive Office of the President, 2016; FSR, 2013; NAS, 2009) attempting to create an environment where methodologies and standards can mitigate risks within forensic bureaus.

Results in this study challenge certain assumptions that have been made in previous research (Dror & Hampikian, 2011; Ask, Rebelius & Granhag, 2008). Firstly, the fact that contextual information has a negative influence on fingerprint examiners' performance. Regarding this topic, it is important to understand that there are

different types of contextual information, and therefore, it is important to observe the effects that each type has within examiners' practice. In this study, individuals' performance was observed by confronting participants with three types of contextual information plus trials where no context was included. Based on previous research, it was expected that individuals would have poorer performance in trials where contextual information was present.

Even though this happened in all trials that presented contextual information, trials within the block that presented contextual information related to the type of the crime showed no significant differences in accuracy when compared with the control block. On the other hand, regarding the time that participants took to complete the trials, it was observed that there were significant differences between trials with contextual information and the control trials. Hence, with the appropriate precautions and limitations, this study allows the forensic community to reflect on the different influences that different types of contextual information have on the decision-making process of fingerprint examiners.

Also related to performance that fingerprint examiners with different ranks of expertise have, previous research has pointed out that individuals with more experience had higher accuracy (Thompson & Tangen, 2014). In this study, it was found that trainees showed lower accuracy levels than examiners and senior examiners. However, when the statistical analysis was conducted, no significant differences were observed. Concerning the response time variable, it was also possible to partially replicate the results from Thompson, Tangen and McCarthy (2014) as trainees seem to be faster than experienced examiners, showing a less conservative approach.

There is also an urgency to understand the need of having a common methodological approach between forensic bureaus regarding the practice of fingerprint analysis. Advocates for the holistic approach argued that the holistic approach is far more approximated to the original rationale that Galton described in his model (Neumann, 2012) and that using a numerical approach is neither scientific nor useful (Champod & Chamberlain, 2013; Champod, 2009; Evett & Williams, 1996). Yet, in many countries, forensic bureaus continue to use a numerical approach contrary to countries such as the UK or the U.S. (Gonçalves, 2017).

Although all research that has been made available seems to claim that the holistic approach is the most appropriate to follow, this study allows some reflection regarding these claims. Results in this study showed no significant differences regarding performance (i.e. accuracy and response time) between fingerprint examiners that follow a numerical or a holistic approach. Despite of what has been mentioned throughout this chapter, there is not the objective to claim which approach (numerical or holistic is most appropriate), however, discussions regarding this topic need to be included in regulators and policymakers' agenda in order to guarantee common practices amongst forensic bureaus in different countries.

The discussion has also been had amongst the forensic community regarding the need for bureaus to be accredited. In certain countries like the UK, the Forensic Science Regulator has required forensic bureaus to be accredited in order to be allowed to provide forensic evidence to courts of law. Similar to the UK, in the U.S. official reports have stated the importance of bureaus working under accreditation standards in order to provide validated evidence. However, participants in this study that worked within non-accredited bureaus showed no significant differences in terms of their performance when compared with individuals that were working in accredited bureaus. These results seem to challenge the accreditation standards and its importance, or perhaps the changes that an accreditation process requires. For a forensic bureau to be accredited it must comply with a number of procedures. Currently there are at least two types of ISO standards that a forensic bureau can become accredited for: ISO 17025 (General requirements for the competence of testing and calibration laboratories) and ISO 17020 (Requirements for the operation of various types of bodies performing inspection). There is also a standard for "Minimizing the risk of human DNA contamination in products used to collect, store and analyze biological material for forensic purposes" (ISO 18385), and currently a new standard has been under development totally dedicated to forensic disciplines (ISO 21043). In the UK the only National Accreditation Body that is allowed to provide ISO accreditation to bureaus is the United Kingdom Accreditation Service (UKAS). Its purpose is to provide an assurance of the competence, impartiality and integrity of forensic laboratories and its examinations and testing activities of forensic evidence. UKAS works closely with the

orientations of the regulator (FSR) in order to promote excellence within processes that are implemented in bureaus in the UK.

One important aspect that bureaus should take into consideration is the fact that even though gaining accreditation could seem to be an exhausting and resource-consuming task, it will definitely have an impact on how laypeople and other professionals (e.g. lawyers and judges) observe forensic practices, preventing potential reputational risks and methodology challenges as observed in some forensic bureaus (Alexander, 2015a).

Results in this study are retrieved only from a single experiment, however, if regulators, academics and policymakers have been advocating for the fact that forensic bureaus need to be accredited, but performance does not seem to show different between accredited and non-accredited bureaus, then perhaps more nuanced considerations need to be given to this topic.

Finally, this study also aimed to observe two variables that were not widely explored previously in the forensics domain, the Need for Cognition and the Motivation that fingerprint examiners have during their daily tasks. Previous research has identified that individuals who have higher levels of intrinsic motivation are less prone to be motivated by external sources and expected to be less biased (Tremblay, Blanchard, Taylor, Pelletier, Villeneuve, 2009). Research also described that individuals who present higher levels of Need for Cognition tend to be more accurate than others with lower levels of Need for Cognition. In this study, it was explored if there were any differences within experts' work characteristics such as level of experience, methodological approach and accreditation of their bureau and these two variables (Intrinsic Motivation and Need for Cognition). Within the three work characteristics mentioned, this study found no significant differences within the levels of intrinsic motivation and Need for Cognition of individuals that (1) had different levels of experience, (2) followed different methodological approaches and (3) worked in bureaus that were accredited or non-accredited. These findings can provide insights that suggest the intrinsic motivation and the Need for Cognition that fingerprint examiners to be not related to these work characteristics but rather to personal traits. Further studies may start to focus more on experts' motivation and Need for Cognition once these

variables have been referred as important factors for individuals' performance (Coelho, Hanel and Wolf, 2018; Hoffman, 2015).

4.6. Conclusion

This chapter explored the influences that different types of contextual information had on fingerprint examiners' accuracy and response time during a task that simulated the Verification phase of the ACE-V process. Participants in the study were based on different forensic bureaus from different countries. In these bureaus, different quality standards were in place as well as following different methodological approaches which enabled to understand results and their implications in a wider perspective. Results in this study demonstrated that research that has stated contextual information to negatively affect examiners' performance can be challenged as well as other official guidelines such as the implementation of accreditation standards or the use of a holistic approach.

5. MOTIVATION OF FINGERPRINT EXAMINERS: ATTITUDES TOWARDS CONTEXTUAL INFORMATION DURING THE VERIFICATION PHASE OF THE ACE-V PROCESS

5.1. Abstract

Fingerprint examiners have been engaging in different experimental activities enabling researchers to achieve a better understanding of their performance regarding the type of influences that variables like contextual information can have. Much has been said regarding the influences that contextual information has within the field of forensics, leading to new guidelines to achieve quality standards within bureaus. Although the human element has been the focus in previous research, it seems that fingerprint examiners' voices have not been actively heard regarding the work they do and the changes that have been occurred within the field of fingerprint analysis. For that reason, this study aimed to explore the opinion of fingerprint examiners regarding influences from contextual information as well as the type of motivation that these individuals have to keep doing their job daily. To achieve that, forty-two fingerprint examiners from fourteen forensic bureaus were interviewed. From the interviews, three main themes were identified. Amongst the most relevant topics, it is possible to differentiate two types of motivation that individuals have when carrying out the job of a fingerprint examiner. This can be useful for recruitment and training actions in the future. It was also described by participants that deficiencies which can influence their performance can be rooted in financial cuts within forensics, which should be addressed by policymakers. Finally, there was no difference regarding the type of motivation that individuals have in relation to the methodological approach that is followed within forensic bureaus or even the type of quality standards that were established at the time of the data collection.

5.2. Introduction

Motivation is a concept that plays an important role within the decision-making processes that individuals carry out in their everyday lives (Jones & George, 2008; Tversky & Kahneman, 1974). Originally, motivation derives from the Latin etymon

movere, that suggests 'action' or 'to move'. However, scholars started to work more broadly within the concept, and new definitions started to become available for the academic community such as Hoffman's (2015) definition who suggested motivation to be "*the degree of effort and intensity toward a goal related to learning or performance*" (p. 8).

Cooper (2002) suggested that individuals' motivation was influenced by personal factors and situational factors, which influences one's performance, observed in variables such as effectiveness, responsibility, autonomy, confidence, creativity and satisfaction. Personal factors, related to features associated with psychological qualities such as disposition, temperament and intelligence, were key to perform tasks with certain sets of skills and abilities, whereas situational factors are found in variables which might not be controlled by individuals such as quality control systems, the size of the organization where one works, the personality of co-workers, types of communication, norms, reward systems and the types of management.

Regarding the influences that individuals experience in their motivation, Ackerman and Beier (2006) found that individuals can feel higher or lower work satisfaction due to the level of challenge they experience within the tasks they are presented with. This work satisfaction is expected to influence the confidence that individuals have when performing the tasks, however, scholars have warned about the dangers that overconfidence might have within one's performance such as cognitive biases (Klayman, 1995; Cohen, 1993; Tversky & Kahneman, 1974; Shanteau & Gaeth, 1983; Golde, 1970).

Another aspect related to the work satisfaction that individuals have is the engagement and enjoyment that they demonstrate when carrying out specific tasks. According to Verplanken and Svenson (1997), people can make three different types of decisions; (1) High-cost decisions (e.g. buying a house), (2) far-reaching consequences decisions (e.g. choosing a career) and (3) significant opinion or emotional value decisions (e.g. voting for a certain political party).

The concept of motivation has been the focus of many studies, allowing the differentiation of aspects within it, such as the identification of intrinsic and

extrinsic motivation (Deci & Ryan, 2000). Intrinsic motivation is observed in individuals who are motivated by doing certain actions, and those actions motivate them to keep doing them as if it was a cycle (Jones & George, 2008). On the other hand, extrinsically motivated people are influenced by external sources such as financial rewards (Shu, 2015).

Regarding their motivation, individuals can present different types of behaviour. Hoffman (2015) described that intrinsically motivated individuals have their personal goals maintained by intrinsic self-improvements. The author identified this type of person as Mastery oriented individuals. On the other hand, Normative oriented individuals are the ones that are extrinsically motivated and who seek motivation from other sources than the stimuli of the tasks they carry out. Thus, this type of motivation is also more prone to be influenced by external factors and result in erroneous decisions based on factors such as cognitive bias (Hoffman, 2015).

Related to fingerprint examiners' opinions, attitudes and emotional states during their work, Charlton, Fraser-Mackenzie and Dror (2011) conducted a study with 13 senior (+7 years of experience) fingerprint examiners. The thematic analysis protocol that was applied to analyse the contents of the interviews described the topics of rewarding, motivation, satisfaction, fear and need for closure as recurring topics throughout the interviews. Within the five categories that the authors described, there were six subcategories related to emotional states and motivational aspects such as the job satisfaction and pride associated with using skills, motivation, satisfaction and hope associated with catching criminals and solving crimes, expression of satisfaction and motivation associated with the importance of the case, feelings towards searching and finding an identification [match] and finally expressions indicating a need for closure on case-work.

Since the results expressed in Charlton, Fraser-Mackenzie and Dror (2011) can be associated with the normative oriented motivation that was described by Hoffman (2015), this study explores these subcategories with a large sample of fingerprint examiners, who had different levels of experience, were from different forensic bureaus, and followed different methodological approaches. During interviews, questions focused mainly on the Verification phase of the ACE-V process.

5.3. Method

5.3.1. Sample

Similar to the study described in Chapter 4, participants were invited to participate in semi-structured interviews through previous contact with gatekeepers who were fingerprint bureau managers. A meeting was arranged by gatekeepers after security checks were passed for the author of this thesis. Fingerprint examiners that volunteered to participate in this study were based in fourteen fingerprint bureaus in nine countries (Brazil, UK, Portugal, U.S., Belgium, China, Australia, Germany, Netherlands). In total 42 fingerprint examiners participated in the study (22 females and 20 males; x_{age} = 38.69 years, s = 11.14, min = 23, max = 60). Participants had on average 9.73 years of experience (s = 9.41, min = 1, max = 35). All participants that were interviewed also carried out the experiment in Chapter 4. Participants were coded with a number and their identity remained confidential.

Fingerprint bureaus where individuals who were interviewed were working followed different types of methodological approaches to conduct their work and had different levels of accreditation. In this study, 23 (55%) fingerprint examiners worked within a numerical approach whereas 19 (45%) followed a holistic approach. Regarding accreditation, 21 (50%) fingerprint examiners worked in bureaus that were accredited either by their national accreditation body or by international bodies of standards such as the ISO standards.

5.3.2. Design

Within this study, semi-structured interviews were carried out with each participant. Interviews took between 45 minutes (min) to 90 minutes (max). Questions that were used as probes were defined prior to the data collection of this study. The decision of using a semi-structured interview instead of a structured interview was due to the fact that this method allowed the author to approach participants with sufficient flexibility to understand individuals' perceptions on the topics related to their emotional states when carrying their work (Bryman, 2012).

Ethical implications were described within the ethics application at the University of Leicester (Appendix F) and forensic bureaus that required submission of an

ethics application prior to data collection. In the ethics applications, it was described that interviews were ideally to be recorded; however, participants did have the right to refuse recordings at any moment and without any consequences for them. In the ethics application, it was also described that personal information could be disclosed throughout the interviewing process. Nevertheless, individuals were always able to comment to the researcher that they did not want a specific part to be recorded. The researcher has experience as a clinical psychologist and was trained to carry out interviews and to provide closure to participants if sensitive information was provided or even if an individual was emotional at any moment. All interview data was anonymised, so participants could not be identified.

The interview purpose and its structure were explained to participants before starting, as well as the use of the audio recording. Participants were asked if they were comfortable with an audio recording of their interview, or if they preferred only written records. Almost all participants agreed to have their interviews audio-recorded, except for one participant who preferred only the author taking notes during the interview. Participants were provided with the option of asking to delete anything that had been said during the interview, however, none of the individuals interviewed asked the author to do so. Each record was given a unique code and stored in a secure location where only the author of the study had access.

Interviews were conducted in different languages by the author of this study. From the total of the interviews conducted, 26 interviews were conducted in English and 5 in Portuguese. Translators were required in 11 interviews (9 Chinese participants and 2 Belgium participants). Translators were native in the language of the participant and fluent in English.

Following Charlton, Fraser-Mackenzie and Dror's (2011) methodology, after the transcriptions, a thematic analysis of the interview content was applied. Thematic analysis is a method that follows pragmatism, which aligned with the mixed methods approach that this thesis followed. This type of analysis was reported by Braun and Clarke (2006) to provide the possibility to organise the data into patterns or themes and to describe it in deep detail. Another possibility that the thematic analysis provides is the triangulation of the qualitative data with quantitative data

that was collected, in order to increase validity within the research carried (Creswell & Plano Clark, 2011).

Three main themes were defined to organise the content of the interviews. These themes were inspired by the Charlton, Fraser-Mackenzie and Dror (2011) study as this was the only study that attempted to observe and identify emotional states of fingerprint examiners. As Braun and Clarke (2006) suggested, a theme is a piece of content that shows a pattern or that has significant meaning within the responses given by individuals. Even though the repetition of contents amongst participants is quite interesting to observe that is not the only way to assign significance to a piece of content (Seal, 2016). The first theme described the practicalities of the work that fingerprint examiners carry out regarding the differences within methodological approaches (i.e. holistic versus numerical approaches) and the accreditation standards in place in fingerprint bureaus. A second theme focused on the influences that contextual information has in fingerprint examiners' performance. This theme was split into two subthemes, one related to the impact of knowing the previous conclusions of other fingerprint examiners and another related to the impact of knowing context associated with the type of crime and the criminal record of a suspect. A third theme described the motivation that fingerprint examiners have to carry out their job as well as negative aspects of their work.

As mentioned in Chapter 2, the results from the interviews in this study were combined with results from Chapter 4. The combination of results with quantitative studies was possible due to the fact that all 42 participants in this study also performed the experiment of Chapter 4. In this chapter, only the contents of the interviews are discussed in relation to the previous literature. In the overall discussion (Chapter 6), the triangulation of data retrieved in this thesis (quantitative and qualitative) was conducted.

5.3.4. Materials

The list of interview questions (Appendix A) was initially developed in consultation with the author's supervisor and the relevant literature and piloted with a pool of fingerprint examiners who were asked if those questions would capture information

on the topics of interest. New questions were added after the results of the second quantitative study (Chapter 4) started to emerge.

Questions were initially developed in English and were subsequently translated into three other languages (French, Portuguese and Chinese) by native speakers. After this first translation, an independent translation also by native speakers was done to identify any differences from the original and the translated versions. Translators completed this work voluntarily and are gratefully acknowledged for their contribution to this thesis.

Meeting rooms where the interviews were carried out were provided by the bureaus where the fingerprint examiners were working at the time of the interview. All rooms were bright, quiet, and had comfortable conditions that allowed the interviews to be conducted without interruptions. All interviews were then transcribed by the author of this thesis using Transcribe open-source software (Wreally, 2019).

5.4. Results

Three themes were defined to organise the contents of the interviews. One theme focused on the differences and common points regarding the Verification phase during the ACE-V process and the relationship with contextual information, a second theme focused on the motivation of fingerprint examiners regarding aspects of their work and its relationship with contextual information and finally, a third theme investigated the influences that system weaknesses promoted in individuals' motivation (single theme without subthemes).

Quotes from the interviews are used throughout this section to illustrate the themes that were defined. Within those, where individuals named colleagues, citations were anonymised to secure the confidentiality agreement that was established prior to the interview. The quotes were not altered in any way such as grammatically, i.e., they were transcribed exactly as the individuals answered. In interviews that were not conducted in English, the author of this thesis translated the speech to English without changing anything from the original quote of the individual.

5.4.1. Theme 1 – Influences on performance related to different methodological approaches and the implementation of different standards of accreditation

Throughout the interview process, participants commented on topics related to the differences between methodological approaches (i.e. numerical approach versus holistic approach) and to the accreditation that fingerprint bureaus currently need.

Participants were split by their type of methodological approach. Amongst the answers from individuals that followed a holistic approach (n = 19), there were four (n = 4) fingerprint examiners (participants 7, 8, 24 and 41) that mentioned that the holistic approach was a better type of methodology in their opinion. Two individuals (participants 7 and 8) of these four started their work following a numerical approach,

> "Back in the time I was so much trained within this numerical thing, I was in the police school so to say for latent prints, and I was bombarded with 'fingerprints are unique! And if you have twelve points in common you know it's him!' and I was already at the time... Are you sure about this? And why twelve? But in the beginning, I didn't say anything because I was just new and the(y) were there for I don't know how many years... So, who am I to say, 'hey maybe that's a bit weird...'. But honestly, looking into the fingerprint with a holistic approach is better than just counting points…" (participant 7)

One individual mentioned that there were not observed significant differences between both types of methodological approaches,

> "I think they're both good. It depends on how you carry out your job. If you do it right, it's not a guideline that will make you better or worse… The same way, if you are not a good examiner, you probably won't be better just because the type of approach you use, right?" (participant 9)

Six individuals (participants 3, 10, 11, 15, 34 and 36) mentioned that although they used to work within a holistic approach, they prefer the numerical approach. Within

their answers, individuals mentioned that their opinion was supported by reasons such as to provide more quality, having more security within their decisions, being more conservative or delivering outputs that are less challenged by courts,

> "We think it could be better if it was points in agreement rather than holistic(…) If I think it's a really poor mark, and not because of the identification, but because of the quality, I would prefer to say I'd like to make that with points of agreement. And that takes you more time." (participant 3)

> "(…) because we don't have a numerical standard, we tend to work to 8 characteristics in agreement" (participant 34)

Within the group of individuals that followed a numerical approach (n = 23), only one individual mentioned that they preferred a holistic approach to be implemented in the bureau where this fingerprint examiner worked, however, the other examiners showed reluctance with that approach,

> "Holistic. However, I think holistic will be hard to implement here. We try to keep the 12 minutiae on the comparison. We try to implement it with the ACE-V, but it's hard with the older ones [referring to older examiners]..." (participant 17)

Equally to the group of fingerprint examiners that followed a holistic approach, there was in this group an individual who mentioned that in their opinion there were no differences between methodological approaches.

> "I do not see any difference. I think that in any place one should find the same results. Obviously that we, because we work with the numerical method, will always report our results with at least twelve minutiae points" (participant 5)

The rest of the examiners in this group (participants 2, 6 and 20) mentioned that they preferred to work using the numerical approach as it provides more security due to the fact that this approach is more conservative.

> “I would say that our method (numerical) is more conservative. It gives more confidence to judges that do not understand much of what we do here” (participant 6)

Besides the conservativeness that the numerical approach provides, one individual mentioned that there are differences within the communication that forensic bureaus have which are related to following different methodological approaches.

> “If a fingermark comes from (country where bureaus mainly work within a holistic approach), and they send it with less than twelve minutiae points, we do not consider that valid. We send it always with a minimum of twelve minutiae points (…) countries are not linked. For instance, here we do not have a technological connection with the PRUM system (…) we have been waiting for it for more than 6 months.” (participant 4)

Regarding the accreditation of bureaus, twenty-one (n = 21) individuals worked in accredited bureaus. Within this group, four fingerprint examiners (n = 4) (participants 9, 10, 11 and 12) stated that even though they thought that accreditation was a good thing to have within the bureau, it was also a factor that added extra work daily. Interestingly, these fingerprint examiners were all working at American bureaus.

> “I'd say that accreditation has made us better, but it always adds that extra work to you. It slows the process a little bit.” (participant 9)

> “Well, it's making a lot more work for us... Because we need to make a lot more of documentation, but I don't think that's necessarily a bad thing. I think it's just... it just changed a lot... Since I'm the technical lead here, I actually write the policies and procedures and so I have to make sure that we follow all of the accreditation procedures. It's been a different experience than when I started. Because when I started it was only ID or NO ID. And it's a totally different mentality when you say it's an exclusion or inconclusive. It is a lot harder, and it's a lot

> more difficult for the examiners, so everything slowed down a little bit. (…) Accreditation made it a lot slower, but I think it's better in the long run because we are able to explain what we do better. I think our training it's better. I think our examiners are better. I think the way we do things is a lot better and more transparent." (participant 12)

Within the bureaus that did not have accreditation at the time of the interviews, fifteen individuals were interviewed. From this pool of fingerprint examiners, only two individuals (participants 36 and 40) mentioned that accreditation seemed to be a good thing. Equally to the group of fingerprint examiners that worked in accredited bureaus, these two individuals claimed that although accreditation was a good thing to have within a bureau, it was also a source of extra work,

> "They are useful, but they have been used in... so for example if I had a case with ten marks in it, I wouldn't be looking to make an analysis and comparison note for every single one of those marks. Whether they were clear marks and clear identifications, or clear negative results. I believe one needs to understand when to do everything, but that's due to management really." (participant 40)

Results relating to this theme shed light on the opinions of fingerprint examiners about the different methodological approaches and the accreditation standards existing within fingerprint bureaus. Regarding the methodological approaches, examiners' opinions showed that a number of participants think the numerical approach is the most appropriate to use since it enables fingerprint comparisons to be more conservative and to secure higher quality. Hence, they felt it is less likely to be challenged by courts of law. Opinions related to the differences of methodological approaches also highlighted the existence of gaps in communication between fingerprint bureaus that follow different methodological approaches. Concerning the accreditation standards, fingerprint examiners were unanimous when saying that accreditation is positive to achieve and hold within a fingerprint bureau. However, a number of examiners (both those working within accredited or non-accredited bureaus) mentioned that all of the work required to achieve accreditation was potentiall problematic from a workload perspective.

5.4.2. Theme 2 – Influences of contextual information on fingerprint examiners' performance

5.4.2.1. Information related to the previous conclusion of the ACE phases by other fingerprint examiners

To observe whether fingerprint examiners would comply with guidelines which have suggested the Verification phase of the ACE-V process to be blindly conducted, this theme related to the information that fingerprint examiners have access to regarding the previous conclusions that colleagues reached prior their verification.

In total, 37 fingerprint examiners mentioned that they have access to the previous conclusions from their colleagues versus only three fingerprint examiners who mentioned that they did not have access to previous conclusions. Two examiners did not have access to previous conclusions, however, these two fingerprint examiners worked in a slightly different setup which is described below this subtheme.

In the group that did have access to previous conclusions, 100% (n = 37) of individuals mentioned that having that kind of information did not negatively affect their performance. On the other hand, also 100% (n = 5) of individuals that did not have access to previous conclusions mentioned that believed they could be influenced by that type of information.

Within the explanations that individuals who have access to previous conclusions gave, one individual provided an opinion mentioning they would be more comfortable if they were able to speak with their colleagues during the verification phase,

> "When I am in doubt if my colleague (referring previous conclusion) is correct or not I feel better if I can call another colleague with more experience" (participant 2)

Regarding discussions between fingerprint examiners with more experience, two individuals (participant 15 and 34) reported that they could change their report when they were challenged by other colleagues with more experience,

> "(…) So I wanted to call it an ID, but my tech reviewer wanted me to call it incomplete but detail in agreement, because he just felt that the better response and the better conclusion we could give was if we could have that tip of the finger rolled up and get a stronger conclusion with stronger weight instead of calling it an ID (…) even though I thought it was an ID (…) I decided to change my decision." (participant 15)

Two individuals (participants 12 and 13) that worked within a non-blind context regarding the previous conclusion mentioned that in their bureau, verifications were seen as challenges to the previous conclusion instead of trying to confirm their colleagues' conclusions,

> "In our verification process, the aim is to disconfirm my previous conclusions." (participant 13)

Regarding the possible disagreement that fingerprint examiners may have within the verification phase, two examiners (participants 13 and 41) mentioned the possibility of attributing the verification to a specific colleague in order to get a higher volume of identifications,

> "You kind of know within the experts who's very conservative... and less... That's not based on ability, that's based on their thresholds. And so, you do get sometimes people shopping for their own identification. That's a very bad practice in my opinion" (participant 41)

Still related to disagreement events, one fingerprint examiner mentioned the need to work collaboratively, i.e., to have discussion throughout the process of the analysis before going to an examiner who carries out the Verification phase,

> "(…) discussions are possible here and actually they are promoted. We believe that discussing a case with a colleague is better when you have doubts" (participant 4)

The same examiner also mentioned the fact that cross-bureau examinations could be a potential solution, however, this is not still possible every time it is needed,

> "Two (examiners) see the mark and they agree, that's fine. When they don't, the mark returns to the first examiner in order to assess it again in a different way. Sometimes it happens that they don't agree in the end. If this happens, we call a third examiner that also needs to sign the final report. If doubts still exist in the end and the third examiner doesn't want to sign the report, there's a guideline that expects us to send it to another bureau outside our office. This was an initiative from ENFSI, however, in this phase, that group is not operational since some countries did show reluctance to be 'evaluated' (participant uses gestures in the word 'evaluated') by other countries. (participant 4)

Concerning deficiencies within the work carried out by fingerprint examiners who were interviewed, one examiner mentioned that the lack of tools and technology affected their way to work and the way they conduct the Verification phase during the ACE-V process,

> "We only have one computer to be used by four examiners and one single magnifier glass. Our cyanoacrylate fuming chamber was handmade by us (…) we do not make verifications as we were supposed to due to the fact that we do not have a system working properly" (participant 1)

One individual also described that the quality of the fingermark could promote the need to talk with other colleagues regarding the previous conclusions they reached,

> "Sometimes if the latent it's not in a good... I do it already. I ask my colleagues what they think. So, if you have to do this alone… There's always a verification... But, yes I ask a lot of opinions. And now I'm a

> trainee and it's better if I can see how other persons look to it than I've done. To find to learn better and to improve." (participant 22)

Finally, an issue that may encourage an examiner to talk to other during the Verification phase of the ACE-V process is the possibility to cut some time off the process as mentioned by two fingerprint examiners (participant 3 and 17),

> "A blind verification consumes you much more time and currently we have a lack of staff and a pipeline that increases every day. We're not machines… yet." (participant 3)

Within the group of participants who mentioned that they do not access the information related to the previous conclusions, there were two examiners (participants 25 and 26) who worked in a fingerprint bureau where the procedure is to access previous conclusions. Hence, they did not access information through their own decision. Their explanation was mainly related to preventing them from being influenced by examiners with more experience,

> "(…) Because of his independence. The preliminary first conclusion, can't influence the second one, no matter how many years of experience the previous examiner has. So, I don't like to have it before finishing my work" (participant 25)

Only one fingerprint examiner seemed to work in a bureau that complied with the standards that require the need to implement blind verifications,

> "We don't receive any information, because we follow a blind verification process. So, let's keep it simple, if one person has been identified within a case, for multiple marks, the first mark that you've identified for that person, will then be put into a package along with a blank technical note form where the verifier can write their conclusions on that sheet. (…) And the only information the verifier can see is the crime reference number". (participant 24)

In the bureau of participant 24, fingerprint examiners can have discussions, however, they prevent the colleagues that they discussed with from being the person that verifies the case,

> "You can have that discussion. Say I'm looking at a mark, I thought it was identifiable but there's maybe a couple of issues due to the clarity, I can go to another examiner and have a discussion with him. Again, I need to write that down. I can do that before it goes to verification. We just need to put a note on the pack to make sure that the person I had a discussion with is not the person that will be the verifier. There's still that openness" (participant 24)

In one bureau where two fingerprint examiners were interviewed (participants 7 and 8), the methodology that was in place differed from all the other methodologies that were reported by the other participants, promoting access to information that was different from other bureaus. In this bureau, fingerprint examiners have access to all information that their colleagues also have. In this specific bureau, the ACE-V process was conducted differently. Instead of having the process with four phases, i.e., Analysis, Comparison, Evaluation and finally Verification, fingerprint examiners in this bureau conducted only the three first phases (ACE) of the process, dismissing the Verification phase due to the fact that in this bureau when a fingerprint analysis needs to be performed, two examiners will conduct the first three phases of the ACE-V process, followed by a discussion between both examiners. Both examiners named this type of methodology as ACE-ACE.

> "Here we do ACE-ACE. If this is the case, the case goes first to [participant's colleague] and then the case will go to me, and we do both the same procedure. Everything equal. (…) When the second one is ready we say, 'hey I finalize the case we can do our discussion today'. Usually, we sit at the table and my colleague or I have all of the pictures of the marks, the analysis phase of the mark and the comparison phase of the mark from each mark, and then we are looking at each other's features because the one who's going to write the report is either one of us and we need to have a consensus of the comparison phase from both." (participant 7)

The other fingerprint examiner in this bureau shared in the interview that they used this type of methodology in order to provide the court with fully transparent reports regarding potential disagreements between examiners.

> "We wanted like this because we know if you do it otherwise, if you don't do your personal notes then the judge will never be able to see what was your process and perhaps the one only found 5 features, the other one has seen 15 features, and of course the one with 5 will say 'oh you found more, let's see what you found', and then we come to agreement on 10 (minutiae features), but then the judge will obviously say 'hey come on, what happened? Because you only found 5 in your analysis phase and then you come with 10 in agreement...' that's not the way we do it. So, to be fully transparent, we both keep our case notes which will be in the case, we will keep the consensus agreement, comparison pictures for each mark and also the final report. (participant 8)

Responses associated with this theme showed that a majority of participants still work within a setting that does not follow guidelines which recommend the need to work within a blind context. This observation was not related either to the type of methodological approach or to the level of accreditation that fingerprint bureaus had since participants that mentioned working within a non-blind context worked either following a holistic or a numerical approach and were from accredited and non-accredited bureaus. Interestingly, participants that worked within a non-blind context mentioned that they were not influenced by knowing the previous conclusion of a colleague. On the other hand, examiners that worked within a blind context mentioned that they preferred to work like that since they expected to be influenced if they were able to access the previous conclusion.

It is also interesting to note that a number of fingerprint examiners that worked within a non-blind context mentioned that they could do a series of actions that are against guidelines such as attributing a case to a colleague to get a higher volume of identifications or to be persuaded to change their report due to others' opinions. Examiners who worked within a non-blind context mentioned that some of the actions conducted were a way to reduce time from the process. Even though

discussions within the process of fingerprint comparisons have been reported to be a good thing to have, participants who worked within blind context mentioned that they had strategies to prevent errors such as attributing the Verification phase to a colleague that they did not speak with about the case or the inclusion of all details related to the discussion being written in the final report in order to show transparency within the process.

5.4.2.2. Influences on performance related to the type of crime in a case (major crimes versus volume crimes) and the criminal record of a suspect

In order to differentiate whether fingerprint examiners thought that information related to the type of crime or the criminal record of a suspect influenced their work, participants' answers were organised in two categories. Firstly, it was assessed whether individuals were able to have access to that type of information, and secondly whether fingerprint examiners believe that this kind of information could influence their decisions.

There were 92% (*n*=39) of fingerprint examiners that mentioned they had access to the type of the crime and the suspect's criminal record during their casework. From this group, 53% (*n*=21) participants mentioned that they do not believe that having that kind of information would influence their performance in any way. Some of the reasons that participants gave regarding their perspective were related to the possibility to increase the quality of their work.

> "Regarding the suspects, if we have their information, we can access the entire ten-print cards, and then observe if we have fingerprints with higher quality" (participant 4)

Two examiners (participant 12 and 6) mentioned that they were not influenced by the type of the crime or the criminals' record, however, there was an increase in motivation when a match was identified in major crime cases.

> "We look for motivation in different ways. If we know it's a homicide or if we have any information, I think that one of the things that

> motivates me (…) trying to find identifications. When you find an identification it's very satisfying and makes you feel special (…)" (participant 12)

Regarding the work carried out when fingerprint examiners know they are working on a major-crime case, one participant mentioned that even though knowing that type of information was not an influence, fingerprint examiners could feel pressure in major crime cases.

> "To some people, it can eventually be a motive of higher attention. For others, it can be a burden of their responsibility, which can also inhibit them. Personally, I am not very apologist of saying 'beware guys that this fingermark comes from a homicide'. I do not think it is going to help. That is my interpretation." (participant 5)

Within the level of attention that fingerprint examiners give to cases depending on their type, two participants (6 and 16) mentioned that even though they did not feel that knowing the type of the crime or the suspect's criminal record would influence them, they probably would provide more time to specific cases.

> "I do not think that (there are influences). If today I have fingermarks from a theft and others from a homicide, I will carry out the homicide first. We have urgency of cases here. (…) If I tell you that it is all the same, I'm lying. If I have on the same day two types of crime, I will dedicate myself to the most important one" (participant 6)

Still, regarding pressures related to the type of crime, one participant mentioned that due to the type of crime, fingerprint examiners can feel pressured to carry out their job faster than usual,

> "Not at all, because latents are latents. Wherever it comes it's a latent. It's only different when we have outside pressures like people or prosecutor asking and saying, it's urgent, you need to do the case. But the work it's the same." (participant 17)

From the group of 39 fingerprint examiners that mentioned they did have access to contextual information related to the type of crime and the suspect's record, 46% of individuals (n=18) claimed that they believed there were influences on their performance due to that kind of contextual information. Amongst the reasons they provided, four participants (3, 4, 40 and 41) mentioned similar reasons to the first group that believed there were influences from contextual information, such as the urgency and the severity of the case that would make examiners allocate more time.

> "The only difference I can think from the top of my head is that for volume crimes, we only go till the 10th candidate, at least on the AFIS system, but for major, we are told to search and research and to be bigger on the search and faster too because it's urgent." (participant 41)

One participant suggested that the influence an examiner can experience when there is access to the contextual information can be related to the sympathy that fingerprint examiners can have with the victim,

> "There is sympathy for the victim, that helps me to do my job better." (participant 25)

Another reason that was provided was related to the context where the crime happened and how it happened. However, this reason seemed not to be related to emotional states but mostly with the practicalities of the analysis of the evidence. Four fingerprint examiners (9, 26, 28, 33) claimed that knowing the context could help them to better understand the surface where the fingermark was deposited and to make their work easier,

> "The information that helps is related to the collection of the print, not to the case type. If I know where the print was, I will do my job better" (participant 26)

> "(…) By analysing the place where the fingerprint was deposited it will be useful for the analysis." (participant 33)

In one bureau where the majority of participants mentioned that contextual information related to the type of the crime and the suspect's criminal record would not influence their performance, two participants (18 and 21) provided a very similar answer. It is important to note that these two fingerprint examiners did have a relationship between them as mentee and mentor,

> "It's important for us to have the information about the case, for the following... to know what happened in the case regarding the evidence only. There is an influence for this, I'm sure, and my mentor did tell me about the effects it has. But for the moment I don't get that influence much because I'm still a young examiner and will have my work always checked by [name of mentor]. It's not a problem... yet (laughs)." (participant 18)

Finally, it is interesting to note that fingerprint examiners enjoy knowing what is happening with the case either during or after the case (participants 2, 13, 36),

> "However, I like to know what happened in the case to know what I am working in" (participant 2)

> "At the end of the day, everybody likes to know what was the case about and how it ended." (participant 13)

Only three fingerprint examiners mentioned that they did not have access to contextual information related to the type of the crime or the criminal record of an alleged suspect. These three participants (7, 8 and 24) mentioned that they believed that this type of information would influence their performance,

> "Obviously there are instances when you know the case needs to be fast-tracked or needs to be processed quickly because perhaps they have an individual in custody. You work in a different way for that because you have a certain pressure for that, and obviously, you start thinking that this case has higher importance than others. When that happens obviously it is due to the context." (participant 24)

Similar to the other group of participants, one participant in this group mentioned that some context could help to carry out the job but only if the contextual information was related to the surface where the evidence was lifted from,

> "I would like to know whether the mark was found on a metal surface or on wood or glass or whatever. Because that will help me interpreting the picture of the mark. Some lines that will be in the mark will be there because the mark was placed on wood, so you got this nerves (ridges) in the wood so I can explain better if there is that difference, because if a line is stopping here and in the reference, print is stopping there I can make an estimation and say what is the reason for that to happen (…) What I necessarily don't want to know is that the metal surface where the print was is a metal surface of an axe or a shotgun, or a knife" (participant 8)

Even though this group of participants mentioned that they do not have any access to contextual information, all three participants (participants 7, 8, 24) shared that they would like to know the context once they did finish their work,

> "Maybe afterwards that's fine, I kind of like to know what happened, what was the crime, to be part of that story…" (participant 7)

Fingerprint examiners had two types of opinions associated with access to contextual information related to the type of crime and the suspect's record. Whilst a number of participants believed that there were no influences associated with access to this information, others thought the opposite. However, the reasons provided by both groups were similar in some key ways. Both groups of participants mentioned that knowing a case was related to a major crime would make them spend more time making the comparison as well as to feel higher pressure to solve the case. Also, both groups mentioned that knowing some information related to the surface where the fingermark was deposited could help them to improve their analysis. Finally, both groups stated that they would like to know what the case was related to in order to motivate them either during the case (for participants working within non-blind context) or after the report was concluded (for participants working within blind context).

5.4.3. Theme 3 – Types of the motivation of fingerprint examiners and influences of negative aspects of the job

5.4.3.1. Sources of the motivation of fingerprint examiners

To the question "what motivates you to carry out your job?", it was possible to differentiate two types of answers. One group of individuals described in their answers arguments that were associated with external sources of motivation such as contextual information (knowing the type of the crime), being involved in the Criminal Justice System, helping society to be a secure place, etc.. Other participants provided answers that relate to an absence of contextual information or others external sources. Instead, in this group, answers related to the particular task of comparing fingerprints. Finally, a third group presented a mixed type of motivation, where both types of arguments were included within their answers.

In the first group, more than half of the participants (n = 24) (1, 3, 9, 10, 12, 14, 16, 17, 18, 22, 23, 24, 25, 26, 28, 29, 30, 32, 33, 34, 35, 37, 38 and 39) provided answers related to external sources that they mentioned would motivate them. Within these answers, four participants (9, 17, 18, and 39) mentioned sources such as feeling that they were part of the criminal justice system.

> "So, I think my contribution to the CJS in general, it's what motivates me. More than just casework really, it motivates me and my job. If I had the same task… if I had just a stack of comparisons and all I can do was comparisons over and over again... There are days where I need to put something aside and go work on something else. And I don't start to feel internal pressures." (participant 9)

Two examiners (25 and 29) mentioned that they were motivated mainly because they were following their relatives' careers.

> "My father was also a police officer. It's like a transgenerational passion. The family values passed through my father to me." (participant 25)

Participant 29 also added that part of their motivation was also promoted by another external source,

> "I also saw some series such as CSI, and I thought I liked this kind of job." (participant 29)

Three participants (12, 16 and 30) mentioned that their motivation was enhanced due to the type of contextual information associated with the type of the crime and by finding matches,

> "(…) here we don't have that information, we look for motivation in different ways. If we knew it was a homicide or if we had any information could be better. I think that one of the things that motivates is trying to find identifications. Because when you find an identification it's very satisfying and makes you feel special if it was a difficult one." (participant 12)

Four participants (3, 14, 23 and 34) mentioned that their motivation was related to the financial reward they would receive to carry out their job (e.g. salary or retirement) and their peer recognition,

> "Well… end of the day, you're doing a job to get paid. So, if they won't pay me I wouldn't be here, but I'm *gonna* be honest with you, I quite enjoy this job. I'm really happy that I found a job that (1) I can do, and (2) I enjoy. General day-to-day, it can be a bit of a boring job at times, it's not CSI Miami, I can tell you that. But, when I do my training, when I do my presentations, or when somebody wants my opinion on something, I like that." (participant 3)

> "I get paid... I think… I want… I want to do a good job. I want my colleagues to appreciate the fact that I am doing a good job. I've always been… doesn't matter what I was doing… I've always been consciously that I wanted to do the best I can. So, that's my personal motivation" (participant 34)

Finally, also related to the financial compensations, one individual shared that there were bonuses for fingerprint examiners who reached a higher number of matches, and that was the source of motivation for this individual.

> "There is a bonus for each match. And the greater the number of matches the better income. The motivation to do the job is getting money." (participant 26)

In the second group of participants in this subtheme, it was observed that 13 fingerprint examiners (4, 5, 7, 8, 13, 15, 19, 21, 27, 31, 40, 41 and 42) described sources of motivation that seemed to be particularly related to the job of fingerprint analysis only.

> "It is a tricky question and I have asked myself that sometimes. In all these years it is not certainly due to the financial incomes, because I do not earn more to identify something. To me it is equal. What gives me satisfaction and enjoyment is knowing that technically I have accomplished. It is the task of identifying that challenges me" (participant 5)

Within this group, two participants (8 and 19) also referred to the task being compared to a puzzle,

> "It's a puzzle! To do the puzzling (…) and I guarantee you it can be hard to look for a very small mark with only 5 or 6 or 7 features in a palm print, but that's for us the most fun we have." (participant 8)

Finally, in the third group, there were seven participants (2, 6, 11, 20, 36 and 42) that described their motivation within a mix between external sources (e.g. rewards, involvement in the criminal justice system, helping society, etc.) and internal sources related to the task (e.g. the task of comparing fingerprints, the puzzle, etc.).

> "It is that satisfaction of being able to compare fingerprints (…) There is a case that was really cool (…) damn, I felt really good. It is my

> golden case, it is that what motivates me, having that job that secures the community" (participant 2)

> "The puzzle… getting hits... that's the buzz, isn't it? it's making the identification right. I don't know 100% if that person has done it, but you know I'm giving the investigation teams some names... I am part of the system (…) passing that information and hearing the reaction of the police officers that we've got a good result... that's motivation." (participant 36)

One participant seemed to have a mixed motivation, however, this appeared to be due to the type of work setup in the bureau where the individual worked.

> "I'm obsessed with fingerprints... I have no idea, I have no idea... People are like why you like so much fingerprints, and I'm like, I just want to do it. I was 17 I knew I wanted to be in fingerprints. I have no idea. And I think the vast kind of... like all the different kind of things we have to do it in (participant's country) (…) I spent the other week a huge amount of time to photograph a horrible print, and we *wanna* search that, we want the reward. I think (going to) the crime scene particularly, finding the fingerprint and being able to search it, and get that reward, it's definitely what motivates me." (participant 41)

Regarding the type of motivation that fingerprint examiners appeared to have, there were three types of motivational factors that participants described. A majority of participants mentioned that they were motivated by external factors which had different sources. Some participants attributed their motivation towards the fact that they felt good to belong to the Criminal Justice System, whereas others stated that they were motivated by the financial reward their work provided. A number of participants also mentioned that they were motivated by knowing what was the crime they were working in. On the other hand, a number of participants mentioned only internal factors as sources of their motivation. Within these factors, participants attributed their motivation to the specific task of comparing fingerprints. Finally, there was a number of participants that showed a mix of external and internal sources of motivation. Amongst these participants, it was

identified factors related to rewards, involvement in the Criminal Justice System and the enjoyment of comparing fingerprints.

5.4.3.2. – Impact of negative issues that affect fingerprint examiners' motivation

Throughout the interview process, 17 participants (1, 2, 4, 6, 7, 8, 11, 12, 15, 16, 18, 19, 20, 23, 25, 26 and 41) mentioned some aspects that they considered negative within their work. Some issues were similar between individuals from different bureaus. Amongst the answers, 15 participants mentioned the lack of investment in the field, particularly in the way training and recruitment were conducted within their bureau as well as the technological or the scientific investment.

> "It is a science that in my viewpoint, is extremely trusted, but we need to prove that is trustworthy. Because that is the question. There should be more international conferences. The community that works here should involve more in scientific projects like yours (mentioning to the researcher). It is not possible to have all people attending, but the bureau should have a section for that (to conduct scientific projects), or to make schedules flexible to the people that do that kind of things, as an incentive. To make us grow and make it a science." (participant 2)

> "I believe that now it is the experts… Technology is… The recruiting processes. And the methodology of the work. Here we work in a different way from our colleagues in other countries, and that is a problem to this field" (participant 4)

> "This field in (participant's country)… I believe that we never did something very specific within the recruitment of specific people." (participant 6)

Besides the small financial investment that fingerprint examiners observed, nine participants (10, 14, 23, 24, 27, 30, 34, 36 and 41) also suggested issues related to

the lack of standardisation between bureaus and the increase of stress due to accreditation processes.

> "Every year there is an accreditation process. There are two types. Every year each department does the evaluation from the (accreditation entity), and then, between three or five years there is an external evaluation. It's very easy the tests they do. It's just a paper. Nothing changes here." (participant 27)

> "So, it's a problem of standards really. And honestly, nobody asked us what we thought about this. You're probably the first one doing this. How are things going to change this way? Probably with more cuts in forensics (sighs)." (participant 36)

> "We really need to move away from those robotic 'you must learn these word for word responses', there's no application for that, so we need to include critical thinking and rational decision-making, and that needs to be started from the beginning with accreditation, but a good accreditation system, not a diploma that you can say you're accredited but continue to do things equally" (participant 41)

Finally, five participants (22, 30, 37, 38 and 40) mentioned that they would appreciate more recognition of their work as well as to have feedback regarding the tasks that they conduct,

> "Knowing the end of the process. Getting feedback. That I would like." (participant 22)

> "Having a bit of appraisal every now and then. Knowing what you've done well and how can you improve… having more activities like this one." (participant 37)

> "Police still see civilian staff as backroom staff. They don't realise that we can sit here and solve the crime for them (lack of recognition). We can probably sit here and solve more crimes than somebody on the

> streets. They see us as clerical staff who can easily be replaced, whose jobs can be done by other people. They don't appreciate the expertise." (participant 38)

Within this theme, participants mentioned negative issues they were concerned with. Amongst those, it was mentioned by a number of participants the lack of financial resources that constrained activities related to the improvement of training or recruitment or improvement within technological assets. The lack of feedback and recognition was also mentioned by participants as well as the extra work attributed to the accreditation process which was associated with stressful moments.

5.5. Discussion

It is important to mention that all the perspectives and opinions that participants shared within the interviews were not judged to be more or less correct. This study intended to observe the different opinions across a wide range of participants regarding the characteristics of their work.

Of interest for this thesis is also the fact that some of the results of this study can be combined with results from Chapter 4 which will be conducted in the following chapter where an overall discussion is presented (Chapter 6). Nevertheless, in this section, results were analysed in terms of the patterns that were observed in each theme and associated with previous literature related to the topics that were discussed across the three themes.

5.5.1. Influences on performance due to different methodological approaches and the implementation of different standards of accreditation

Commencing with the first theme related to the different methodological approaches and standards of accreditation, it is important to observe that within the sample of bureaus where fingerprint examiners that participated in the interview process were working, there was a balanced number of accredited and non-accredited bureaus which were either following a numerical or a holistic approach.

Official guidelines in these countries do not oblige bureaus to follow a holistic approach in order to be accredited. Some regulators suggest fingerprint laboratories should relax the requirement of having a minimum number of minutiae, and instead have a greater focus on the quality of each comparison (FSR, 2019; 2013; 2011). Despite the fact that in most countries, the numerical approach was first used (Neumann, 2012), entities such as the IAI mentioned in the past that the numerical approach was scientifically poorer than the holistic approach (Champod & Chamberlain, 2009).

Anglo-Saxon countries like the UK or the U.S. have had their standard challenged in the past (Evett & Williams, 1996). However, in certain countries, a numerical approach still is the standard that is followed within the practice of fingerprint analysis such as Portugal where the standard requires 12 minutiae points in accordance plus an extra point named "security point" or Germany, where fingerprint bureaus follow a 12-points standard. In Italy for instance, fingerprint examiners need to reach a minimum of 16 minutiae points in order to declare a comparison as a match.

Within the reasons mentioned by participants in this study, it was observed that conservativeness within the process of fingerprint analysis was highlighted which seems to go against what has been suggested by different authors regarding abandoning the numerical approach (Champod & Chamberlain, 2009; Evett & Williams, 1996). What seems to be of interest is the fact that the numerical approach, or the notion of counting minutiae points, still is a practice that fingerprint examiners do whether they work within a numerical or a holistic approach.

Concerning the need to be conservative, there might be an association with what Charlton, Fraser-Mackenzie and Dror (2010) mentioned in their theme of "need for closure", as being conservative may allow fingerprint examiners to make themselves distant from the case and able to close a case. In this study, a number of examiners mentioned that the numerical approach would be more suitable to present as evidence in court. These results are in line with what previous literature demonstrated when describing fingerprint examiners who act more conservatively to prevent innocent people from being wrongfully convicted (Tangen, Thompson & McCarthy, 2011).

Regarding the need to present evidence at court and reducing the likelihood of facing challenges, Mnookin (2010) mentioned that courts and jurors require "*in many cases, as a minimum prerequisite to admissibility is simply much better error rate information about examiners' abilities in practice*" (p. 1243). Hence, it seems sensible to expect fingerprint examiners to be more comfortable with presenting evidence that seems to be more conservative in order to be less challenged in pressured environments such as a court of law, where evidence from forensic sciences has been questioned (Pyrek, 2007) and challenged by public jurors as demonstrated in countries like the U.S. (Kaplan, Ling & Cuellar, 2020) or Australia (Ribeiro, Tangen & McKimmie, 2019).

Regarding the accreditation standards and the opinions that fingerprint examiners had regarding that topic, a small number of fingerprint examiners that either worked in accredited or non-accredited bureaus mentioned that having accreditation was a seen as a good thing. An interesting matter to highlight was the fact that individuals from both types of bureaus felt that accreditation was a practice that would add extra efforts within casework.

Even though accreditation has been suggested by official entities such as The National Commission on Forensic Sciences (DOJ & NIST, 2016) and the Forensic Science Regulator (2011), it is important to observe the difficulties that public bureaus face when accreditation processes are implemented as well as the workforce needed to comply with all the requirements that an ISO standard entails. In the field of DNA for instance, Peterson and colleagues (2003) observed that an important feature of an accredited bureau, such as implementing blind proficiency testing, was easier to conduct in private labs where the turnaround times are shorter than in public labs.

The third theme which described negative aspects of examiners' work, observed opinions related to the accreditation process which instead of being referred to as a positive feature to have within a forensic bureau, has been described as a stressor for individuals' where the extra work mentioned by a number of examiners can also be related to a potential time pressure that has been referenced by other examiners that were interviewed in previous research (Charlton, Fraser-Mackenzie & Dror, 2013).

5.5.2. Access to contextual information during the Verification phase of the ACE-V process

In these subsections, individuals' responses regarding their access to contextual information were analysed as well as how labs comply with official guidelines. Fingerprint examiners' performance was widely focused in the subsections, however, the combination between what fingerprint examiners mentioned and their accuracy in tasks that were conducted in this thesis, namely in Chapter 4, are discussed in the overall discussion in Chapter 6.

5.5.2.1. Access to previous conclusions by other fingerprint examiners

In the second theme related to the access to contextual information that fingerprint examiners have, it is important to remember that official guidelines within the field of fingerprint analysis stated in the past that fingerprint bureaus should consider within their procedures to have blind verifications in order to mitigate the risk of cognitive bias as other fields (e.g. biomedicine) usually do (PCAST, 2016; Forensic Science Regulator, 2015).

Interestingly, 88% (n = 37) of the participants that were interviewed had access to the previous conclusions of their colleagues, even though 50% (n = 21) of participants worked in accredited bureaus. It was also interesting to observe that 100% of individuals that had access to contextual information related to previous conclusions mentioned that they did not feel they were negatively affected by this in their performance. This issue can be related to what Shanteau and Gaeth (1983) mentioned regarding overconfidence that experts may experience, or even the lack of awareness regarding the effects of contextual bias within the field of forensics (Kassin, Dror & Kukucka, 2013)

There was a pattern related to the experience level of fingerprint examiners in moments of disagreement, which could influence final decisions of examiners that were less experienced. These results may suggest that there is a gap for potential bias such as the confirmation bias (Edmond, Tangen, Searston & Dror, 2015; Nickerson, 1998; Tversky & Kahnemann, 1974), where individuals may anchor in previous beliefs as was described in previous literature (Bonefeld & Dickhauser,

2018; Darley & Gross, 1983) or be influenced by experience and confidence levels (Cohen, 1993; Shanteau & Gaeth, 1983).

Another point of interest was the fact that a number of interviews heard that discussions between examiners during the procedure of fingerprint analysis seemed to be a beneficial strategy, however, in only one interview the procedures seemed to comply with what standards recommended regarding discussing with colleagues and not having those specific individuals conducting the verification phase of the work (participant 24). Concerning the possibility of having multiple opinions, Liedtka (2015) mentioned that this strategy could be applied to decrease potential flaws within a decision-making process. However, only two fingerprint examiners that worked in the same bureau described that discussions during the Verification phase in their bureaus were seen as a challenge of prior conclusions as Liedtka (2015) suggested.

Blind analysis has been widespread in scientific communities (e.g. physics, social sciences) as a strategy that promotes the observation of results without the influence of previous information (Roodman, 2003). MacCoun and Perlmutter (2015) described the blind analysis as "an optimal way to reduce or eliminate experimenter bias and the unintended biasing of a result in a particular direction" (p. 1). It is possible to implement blind analysis in different ways, such as scrambling labels of data, asking other colleagues to assess the same information without knowing what they are assessing or by arranging a procedure where the same individual assesses the same data without knowing it (MacCoun & Perlmutter, 2015).

Another issue that seemed to be contrary to what previous literature mentioned (PCAST Report, 2016; Kassin, Dror & Kukucka, 2013) was the fact that fingerprint examiners were able to choose colleagues to make verifications, instead of working within a linear sequence and preventing themselves to access irrelevant information that could promote biased decisions within their workload.

Although 100% of participants that had access to previous conclusions felt that their performance was not affected negatively by it, a point of interest was raised by some examiners regarding the lack of tools to carry out their work during the Verification phase. Gibb (2019) mentioned that the lack of financial resources to acquire

materials and human resources has been putting practice at risk. Related to this topic, in Theme 3, a number of fingerprint examiners also mentioned that a negative thing that they observed in the field of fingerprint analysis was the lack of resources and the pressures to finish their work. Related to the lack of resources, fingerprint examiners also mentioned that working within a non-blind context towards previous conclusions made completing their work faster.

In this subtheme, at least two main issues were observed. In order to promote best practices within the Verification phase regarding the access to previous conclusions, it is important to demonstrate to fingerprint examiners that contextual bias associated with access to previous conclusions is a reality as demonstrated by previous literature (OSAC, 2015; Kassin, Dror & Kukucka, 2013; Dror, 2013; Dror, 2012; Langenburg, Champod & Wertheim, 2009). However, policy and decision-makers need to be aware that fingerprint bureaus may need more resources to carry out their work properly.

5.5.2.2. Access to information related to the type of the crime or the criminal record of a suspect

When analysing answers regarding access to contextual information related to the type of the crime or the criminal record of a suspect, it was observed that only 7% (n = 3) of participants did not have any access to that type of contextual information. All of these fingerprint examiners mentioned that they believed that kind of information would influence their performance which is in line with literature that observed the effects of contextual information in forensic disciplines (Earaker, Morgan, Harris & Hall, 2015) as well as in other fields (Veletsianos, 2010).

Even though literature in the field of cognitive psychology has demonstrated that contextual information can be a cause of flawed decisions, here it is also discussed that having some types of contextual information are positive factors for performance. McRobert et al. (2013) found that contextual information had a positive influence on individuals' decisions. In the study conducted in this thesis, individuals had different views regarding access to contextual information related to the type of crime. Whereas some individuals mentioned that they wanted to know

what happened in the case in order to better understand details related to the evidence, such as knowing the surface where the fingermark was deposited, other examiners mentioned that they would be more motivated by having that kind of context.

Regarding the assumption that some types of contextual information can have a positive influence on fingerprint examiners' performance, two points should be raised. Firstly, it is important to address the literature that has shown contextual information such as details of the case to have an influence on forensic examiners' accuracy as demonstrated in the field of DNA analysis by Ask, Rebelius and Granhag (2008) or in the field of bitemarks by Osborne, Wood, Kieser and Zajac (2014). However, it is also important to note that fingerprint examiners in this study identified contextual information to be either a source of motivation as these individuals like to know what the case was about, or to retrieve some details related to the fingerprint analysis process. Whereas knowing information about the case was a topic widely referred to in the theme related to negative aspects of the job (see below references related to the lack of feedback), it is also important that only one bureau where two fingerprint examiners were interviewed had a group of scientists within their team that filtered information throughout the ACE-V process.

Still relevant to the fact that a high percentage of participants mentioned that they liked to know the details of a case that they were working on, even though bureaus did not allow this information for fingerprint examiners to access, management systems may need to consider changes within some of their procedures since the workflow seemed to be directly affected by the type of case as a number of examiners mentioned that they would need to work faster in high profile cases. Thus, besides the perspective that examiners may have regarding the effects of contextual information and the lack of investment within bureaus, it is important to highlight that systems that manage cases seem to not comply with official guidelines regarding access to contextual information.

5.5.3. Different sources of motivation within fingerprint examiners' work and negative aspects of casework

5.5.3.1. Different sources of motivation within fingerprint examiners' work

Two patterns emerged in how fingerprint examiners described their motivation to carry out their work. The first type of motivation referred to external sources such as the contextual information within cases, the involvement with the Criminal Justice System or their role to help society to be secure. Patterns related to external sources of motivation were also found in previous literature (Charlton, Fraser-Mackenzie & Dror, 2013). These types of motivation are extremely important to address since they can be related to potential flaws within decision making in highly covered cases such as the Brandon Mayfield Case (Thompson, 2005) or literature that claimed that contextual information influences individuals' performance, in a variety of domains (McRobert et al., 2013; Rovira et al., 2013; Veletsianos, 2010; Clarke et al., 2000) as well as in forensic sciences (Osborne et al., 2014; Kassin, Dror & Kukucka, 2013; Dror & Hampikian, 2011; Langenburg, Champod & Wertheim, 2009).

Aiming to continue the work of Charlton, Fraser-Mackenzie and Dror (2013), this study defined the types of motivation that fingerprint examiners have. In the other part of the group, participants mentioned that they were motivated to carry out their job by factors that were associated with internal sources of motivation such as enjoying the task itself. This type of motivation seems to be more inclined to guidelines that mention a blind context as these individuals seem not to need anything besides carrying out the task of fingerprint analysis to be motivated.

Jones and George (2008) found that motivation was a central factor for decision-making processes and to better understand individuals' behaviour. According to Tversky and Kahneman (1974), the motivation that one has will play an essential role in decision-making as well as strategy choice. This seemed to be the case in this study since the participants differed in the type of strategies they used in order to maintain their motivation.

Regarding the different types of motivation that individuals in this study showed, both are in line with what Hoffman (2015) mentioned concerning the two types of motivation that can be observed, namely the extrinsic motivation and the intrinsic motivation. Extrinsic motivation, where Normative Oriented people are included, has been described as being promoted by external sources such as rewards or other types of recognition. In this study, individuals that mentioned external sources of motivation can be placed in this category. On the other hand, intrinsic motivation, where Mastery Oriented individuals are placed, increases due to the learning process and by being cognitively active, which is in agreement with individuals in this study that mentioned that they were motivated by the task itself.

The results found within this topic of examiners' motivation is of interest to the fingerprint community since it provides insights regarding examiners' motivation. Ideally, one could suggest that all fingerprint examiners should be more intrinsically motivated since this type of motivation has been described as less prone to external influences and therefore less expected to commit errors due to contextual information. However, fingerprint examiners are not machines and have their personality traits that shape factors like motivation. However, this new knowledge is important because it (i) acknowledges that examiners are motivated by different factors, and (ii) opens the discussion regarding new strategies to motivate individuals such as new methods of training or perhaps management strategies that enhance activities such as regular feedback or that provide insights related to communicating the outcome of cases which has been explored in previous literature (Earwaker, Morgan, Harris & Hall, 2015)

5.5.3.2. Negative aspects of fingerprint examiners' job

Although performance in terms of accuracy and response time and the motivation of individuals are explored in the overall discussion in Chapter 6, it is important to acknowledge that the motivation one has depends on personal and situational factors (Cooper, 2002). Whereas personal factors are observed within each person's personality (e.g. disposition, temperament and intelligence that promote individuals' abilities and specific skills), situational factors can be out of one's control, and therefore can become negative aspects of the job. These aspects of the

job are analysed here in order to discuss whether they could influence fingerprint examiners' performance.

Three topics were highlighted from participants' answers. The first topic was related to the lack of investment within fingerprint bureaus. This seems to be of high importance as it has been observed that the lack of financing has been putting the quality of the work of forensic bureaus at risk (Gibb, 2019). Regarding the issue of funding cuts, the Forensic Science Regulator mentioned in the past that the continuation of cuts within forensics in the UK has been contributing to a series of potential problems such as the lack of time for scientists to prepare reports for court as well as the incapacity to achieve quality standards (FSR, 2018). It is plausible to think that this lack of funding can also influence how fingerprint examiners see the issue of accreditation standards. When bureaus appear to be struggling to run efficiently, it seems that extra work within the same conditions will decrease fingerprint examiners' motivation, even though a number of individuals mentioned that being accredited was a good feature for a bureau to have (subsection 5.5.1.).

It appears obvious that lack of investment can be a source of tension and pressures such as time constraints, which puts the quality of the work at risk as previous research has shown (Huber & Kunz, 2007; Maule & Edland, 1997). In the UK for instance, the FSR has provided evidence to the House of Lords arguing to increase funding to promote best practices (FSR, 2019), however, numbers have shown a decrease in terms of available funding for forensic disciplines since 2008, where the funding line was reduced from £120m to an estimated £50m in 2018 (Cookson, 2019).

With all of the funding issues mentioned above, it seems reasonable that fingerprint examiners mentioned the arguments observed in this study regarding the lack of resources and the potential to improve their methodologies with ongoing training. Regarding this matter, it seems that policymakers should be aware of forensic bureaus' needs and perhaps should advocate for greater support of regulators such as the FSR and its other counterparts in other jurisdictions.

Finally, an issue related to the communication and recognition within bureaus and between different departments was addressed. In line with the study by Charlton,

Fraser-Mackenzie and Dror (2013), a number of participants mentioned the importance of feedback regarding their work. The value of providing feedback has been widely demonstrated (Boud & Molloy, 2013; Croskerry, Singhal and Mamede, 2013; Kerstholt & Raaijmakers, 1997) as a strategy to test accuracy and allow for adjustments that are needed in a specific time. This being said, it seems that implementing a feedback system with the right conditions would promote not only the motivation of fingerprint examiners but also the possibility for them to continuously improve their performance. This type of implementation could, perhaps, be implemented as follow-up step after the case is closed. Bureaus could implement a system where internal audit teams would review cases and assess proficiency within a totally blind context and compare results with the ones reached during case-work. Results could then be utilised either for training and/or research purposes. In the event of reaching a conclusion that putted in risk the decision that was reached, an investigation should be opened in order to assess if a different report should be conducted. This type of approach can put in risk forensic bureaus' work, however, the mindset that should be used for this type of procedure is the improvement of performance and not to point out errors of examiners.

Besides the adjustments and the possibility to improve performance, it seems that in the answers of participants in this study, the need to have feedback was also associated with the desire for recognition and to feel part of their departments. Regarding this point, it seems that management should promote conversations or informal gatherings between professionals from different areas in order to increase inclusiveness as well as providing a better environment to work in. Also related to the first topic of this subsection, another alternative could be to send fingerprint examiners for international training in order to learn how other fingerprint bureaus work and to stimulate conversations across different agencies aiming to create a greater sense of community within the fingerprints domain.

5.6. Conclusion

In this chapter, a qualitative study was conducted where 42 fingerprint examiners were interviewed. Three topics were discussed during the interviews. Those were related to the differences between methodological approaches and quality standards that are followed in bureaus based in different countries, and the effects on

performance influenced by contextual information. From the opinions that participants shared during the interviews, it was possible to have a wide view regarding fingerprint examiners' opinions towards these subjects as well as to other features of their work such as negative issues they would like to see addressed. Moreover, the results of this study shed light on some of the results observed in Chapter 4 which will be discussed together in the following chapter.

6. OVERALL DISCUSSION

6.1. Introduction

This chapter aims to combine the results from the three studies that were conducted in this thesis. The quantitative results from the experiments showed two distinct aspects. Firstly, the online experiment (Chapter 3) presented the differences in performance between fingerprint examiners and laypeople in relation to the influences of contextual information on their accuracy and response time. It could be observed in Chapter 3 that experts were significantly slower than laypeople during the pattern recognition tasks. However, when observing accuracy, experts performed significantly better than laypeople. These results are in accordance literature that claimed that experts will have their abilities better developed for tasks that they have familiarity (Weiss & Shanteau, 2014). Regarding the effects of different contextual information in individuals' performance, results in Chapter 3 are partially in accordance to previous claims that stated contextual information to negatively affect performance (Alexander, 2015a; Alexander, 2015b; Osborne, Wood, Kieser and Zajac, 2014; Kassin, Dror & Kukucka, 2013; Dror & Hampikian, 2011; Tangen, Thompson & McCarthy, 2011; Garret & Neufeld 2009). There was one contextual information variable that did not have a significant effect on accuracy neither on experts nor laypeople which was the type of crime/publication. However, all types of contextual information made both types of participants to be slower making decisions, providing an insight regarding the effect it had on their time response. These insights that provide the assumption that some types of contextual information does not influence accuracy can be a first step to challenge

the notion that all types of contextual information affect individuals' (either experts or laypeople) performance.

The experiment that was conducted on-site in Chapter 4 contributed to better understand and gain more knowledge about fingerprint examiners' performance and the effects that different sources of contextual information have on tasks that mimicked some of their daily work, specifically during the verification phase of the ACE-V process. Due to the multiplicity of laboratories that participated in Chapter 4, it was also possible to observe how different procedures and quality standards that are established in different bureaus can influence fingerprint examiners' performance. Even though previous literature claimed that contextual information can influence performance during fingerprint examination (Earwaker, Morgan, Harris & Hall, 2015; Dror, Kassin & Kukucka, 2013; Langenburg, Champod & Wertheim, 2009), results in Chapter 4 challenged this notion that all types of contextual information have a negative impact on individuals' performance as throughout the chapter, it was observed that one of the variables ("type of crime") did not affect significantly participants' performance. Furthermore, within the study in Chapter 4 two variables related to the motivation and the need for cognition of individuals were explored. These two variables have been explored in previous studies. Regarding motivation, Tremblay, Blanchard, Taylor, Pelletier, Villeneuve (2009) observed that individuals with higher levels of intrinsic motivation will be influenced by internal motivator factors, therefore less prone to be influenced by external sources such as motivation by contextual information. In relation to the Need of Cognition that individuals present when carrying a task, research showed that higher levels of Need for Cognition tend to be promote accuracy. Within study 4 three work characteristics were defined (levels of experience, methodological approaches and bureaus' accreditation). Participants showed no significant differences within their Intrinsic Motivation and Need for Cognition regarding any of the work characteristics. These findings provided insights that can suggest the need to reflect more on personal traits of examiners rather than work characteristics, hence additional studies should focus more on experts' motivation and Need for Cognition once these variables have been referred as important factors for individuals' performance (Coelho, Hanel and Wolf, 2018; Hoffman, 2015).

6.2. Influences on performance related to the methodological approaches and the quality standards within fingerprint bureaus

According to previous literature and guidelines (Champod & Chamberlain, 2013; Neumann, 2012; Evett & Williams, 1996), it was expected that participants who worked in accredited bureaus and followed a holistic approach would be more efficient regarding their accuracy and response time when conducting their work, and therefore when carrying out tasks such as the trials within the experiment in Chapter 4. Interestingly, that was not the case. Results showed no significant differences between the two methodological approaches or accreditation status with regards to accuracy or response times. Regarding this observation, one can reflect the importance of these two features within fingerprint analysis work.

The design of this thesis makes it possible to reflect on and combine results from Chapter 4 within the context of the interviews (in Chapter 5) in which fingerprint examiners reflected on their opinions related to using a holistic approach and the need to be accredited. Regarding the holistic approach, fingerprint examiners mentioned that when following this approach, they continue to '*count minutiae points*' since this is still considered important for a number of reasons such as '*not being challenged at a court of law*'. The views about the use of a holistic approach may explain why there were no observed differences in performance between both methodological approaches in the experiment. It suggests that fingerprint examiners keep using numerical standards within the holistic approach. Therefore, it seems that the literature advocating the use of a holistic rather than a numerical approach may need to be reassessed, in order to observe whether (i) this methodological approach is, in fact, the best approach for fingerprint examiners, (ii) what is the best way to promote new procedures within bureaus, and (iii) how to empower fingerprint bureaus and fingerprint examiners to be an active piece in promoting similar practices and communication between bureaus in different countries.

Furthermore, the differences of methodological approaches and the choices of following either the holistic approach or the numerical approach that were observed in this thesis seem to be in line with the argument that a standard that is robust enough to be used across forensic bureaus is needed. One potential solution to this challenge is the use of likelihood ratios within the task of fingerprint analysis (SWGFAST, 2011; Champod, 2009), which can be achieved by promoting internal

research within laboratories as some individuals mentioned during interviews in Chapter 5. Likelihood ratios are more objective when reporting evidence, however, this type of information is also more difficult to interpret by laypeople (Thompson & Newman, 2015) It is important to note that this solution would be in line with the work that is needed to be able to use likelihood ratios (Kahneman & Tversky, 1973) (i.e promoting internal research).

The accreditation standard that is expected to be implemented in forensic bureaus was acknowledged by fingerprint examiners regarding its value, however, a number of individuals also mentioned that this requirement could also be seen as a negative feature that would only create '*extra work*', and that it could be seen only as '*a diploma for the wall*'. Consequently, it is of interest for the forensic community as well as for the regulators to reflect on two points. Firstly, the fact that accreditation should be promoted within forensic bureaus in a positive way instead of being seen as only extra work which requires financial investment as mentioned by Smith (2019). Secondly, the fact that results in Chapter 4 suggest that individuals who work within accredited bureaus did not excel in performance when compared with their peers that work in non-accredited bureaus, suggests regulators should give thought to the accreditation process itself, the real changes that this feature promotes within bureaus and mainly in individuals' methodologies.

During the trials in Chapter 4, it was also observed that individuals working within a holistic approach were slower than those who followed a numerical approach regardless of whether contextual information was presented or not. Furthermore, individuals working in accredited bureaus were also slower than individuals from non-accredited bureaus when confronted with previous conclusions made by other individuals. Besides taking more time, these results may suggest individuals were more conservative (e.g. taking more time to evaluate ridge details) since they took longer to make a decision instead of being impulsive and faster which is in line with what previous research mentioned when describing fingerprint examiners who seemed to be more conservative than their peers with less experience and less training (Thompson, Tangen & McCarthy, 2014; Thompson, Tangen & Searston, 2014). However, this only relates to the time that participants took to complete the task and not necessarily to their accuracy. Hence, it may be that the methodological approach and the quality standards that have been advocated by academics and

regulators only make fingerprint examiners slower, without being necessarily more accurate. However, it is important to highlight that fingerprint examiners acknowledged the importance of accreditation, and aim to provide high-quality comparisons, even though, further research should be encouraged in this area to determine the relationships between methodological approaches, accreditation and fingerprint examiners' performance.

6.3. Considerations regarding the access to contextual information and its influences on fingerprint analysis and examiners' motivation and Need for Cognition

Within the discipline of forensic science (Earwaker, Morgan, Harris & Hall, 2015; Edmond, Tangen, Searston & Dror, 2015; Osborne, Wood, Kieser & Zajac, 2014; Kassin, Dror & Kukucka, 2013; Dror, Wetheim, Fraser-Mackenzie & Walajtys, 2012; Dror & Hampikian, 2011; Langenburg, Champod & Wertheim, 2009; Ask, Rebelius & Granhag, 2008; Hall & Player, 2008), as well as in other fields (Bonefeld & Dickhauser, 2018; McRobert et al., 2013; Veletsianos, 2010; Munro & Stansbury, 2009; Clarke et al., 2000; Egglin & Feinstein, 1996; Klayman, 1995; Darley & Gross, 1983; Zukier, 1982), the effects of cognitive bias have been discussed at length. In all of the studies, a common conclusion has been reached, that cognitive bias can be associated with the presence of contextual information within experts' work. However, previous studies have failed to address two points. Firstly, studies focusing on cognitive bias within forensic sciences seem to have mainly focused on experts' errors, instead of identifying if those errors are also observed in a general population of laypeople. Secondly, studies have failed to observe the different effects that distinct types of contextual information have within individuals' decision-making process. Furthermore, previous research has not paid attention to observing factors that are important for individuals' performance such as motivation and Need for Cognition.

In the online study conducted in this thesis (Chapter 3), it was observed that even when confronted with contextual information, fingerprint examiners had higher accuracy than laypeople when analysing a stimulus that mimicked fingerprints. In this study, both experts and laypeople suffered from the effects of contextual information and cognitive bias when comparing stimuli familiar to both groups (i.e.

the excerpts of text). For both types of stimuli (artificial fingerprints and excerpts of text) fingerprint examiners had higher accuracy and lower response times than laypeople, highlighting experts' skills. Secondly, in both groups of participants, an interesting fact occurred. In both types of stimuli laypeople and fingerprint examiners were less accurate when confronted with contextual information related to the record (criminal or academic) of another person (suspect or student) as well as to the previous conclusion made by another person.

It is also interesting to point out that one type of contextual information did not have significant effects in participants' performance. Both groups of participants in study of Chapter 3 showed no significant differences in their accuracy and response times in trials where contextual information related to the type of the crime (when using artificial fingerprints) or the type of publication (when using excerpts of text) was presented when compared with the control trials where no contextual information was presented. This observation is in line with Hall and Player (2008) which demonstrated that fingerprint examiners did not suffer when confronted with contextual information related to crimes with different severity. What is novel in the study conducted in Chapter 3 is the fact that laypeople and experts showed no significant differences, which suggests that this type of contextual information may not be a negative influence for fingerprint examiners' work.

Regarding the effects that contextual information is known to have on fingerprint examiners' performance, the second study in this thesis explored in more detail the results from the online study in (Chapter 3). In chapter 4, similar results regarding the effects of contextual information were found in individuals' performance, specifically in their accuracy and response time. Responding to some of the limitations in previous research where participants were from one single bureau (Thompson, Tangen & McCarthy, 2011) or to designs that only used one type of contextual information (Hall & Player, 2008), participants in this study showed no significant differences regarding their accuracy and response time when carrying out comparison tasks in the presence of contextual information related to the type of crime. In this study, it was also observed that the type of contextual information that affected participants' accuracy and response time was the information related to previous conclusions from another examiner.

These findings provide an opportunity for discussing whether all contextual information increases the chance of flawed decisions. The qualitative study in Chapter 5 shed some light on this discussion. During interviews, a number of participants mentioned that accessing contextual information related to the type of crime was something that they find beneficial since it enhances their motivation and interest in their work. This argument was mentioned not only by examiners that worked within a non-blind context who are used to having access to contextual information but also by fingerprint examiners that worked within a blind context. As a matter of fact, in the qualitative study, it was observed that a number of participants had what Hoffman (2015) described as a Normative Oriented motivation, i.e. individuals whose motivation is rooted in external variables such as the context. Hence, it seems that some fingerprint examiners are not exclusively interested in the work of fingerprint analysis due to the task itself, but rather because of other factors that motivate them such as being engaged in the criminal justice system, helping society to be secure, identifying criminals, as well as knowing the type of the crime. Hence, one may infer that context related to the type of crime may be beneficial to performance rather than problematic.

Although research and guidelines that advocate to work within a setting without any relation with contextual information are important to acknowledge (DOJ & NIST, 2016; PCAST Report, 2016; FSR, 2015; MacCoun & Perlmutter, 2015; Champod, 2014; Kassin et al., 2013), one cannot ignore what fingerprint examiners feel and think about the setting that they are asked to work in. Regarding this, three points deserve further attention.

Firstly, the fact that some sources of contextual information, which do not seem to have any impact in individuals' motivation and that seem to make them reduce performance (e.g. knowing the previous conclusion) have been made available within bureaus, whereas other types of contextual information, were mentioned as a motivating factor and have not been available to examiners. Having access to previous conclusions may accelerate casework as mentioned by some fingerprint examiners, however, it seems inefficient to provide contextual information that has no meaning for examiners' motivation as it makes them have poorer performance just to decrease turnaround times.

Secondly, the fact that isolating fingerprint examiners from some contextual information that is related to the cases, even after the case is closed, can be a source of demotivation and loss of interest as some examiners felt that the lack of feedback was a negative feature they would like to change within bureaus and system's management. Research (Boud & Molloy, 2013; Croskerry, Singhal & Mamede, 2013; Bahrami et al., 2012; Kahnemann, 2011a; Bishara & Busemeyer, 2008) has found that feedback is a way to make individuals increase their performance as well as to introduce new procedures. Within the field of fingermarks submission, laboratory practitioners seemed to benefit from feedback, although this strategy has not been widely implemented (Earwaker, Morgan, Harris & Hall, 2015). It seems that the inclusion of feedback could be a tool used to introduce new practices such as proficiency testing with regular feedback within bureaus, promoting good practices and using a strategy that motivates fingerprint examiners.

Finally, still related to the access to contextual information and its influence on individuals' motivation, it is plausible to think that a recruitment matter also needs to be discussed. A number of fingerprint examiners mentioned during the interviews that recruitment and training was a negative issue. However, if the standardisation of a common practice only retrieves insights from academics and policymakers rather than including fingerprint examiners within the discussion, changes may become more difficult to achieve. Potential starting points could be (i) to include more often fingerprint examiners engaged in discussions, (ii) to empower them to be part of the process, and part of the change which may unify bureaus across countries and procedures worldwide and (iii) to improve decision-models within the practice as Earwaker, Nakhaeizadeh, Smit and Morgan (2020) have recently suggested. Recently created, the Forensic Capability Network (FCN) aims to "*deliver high quality, specialist forensic capabilities in support of the NPCC's 2025 policing vision to rapidly protect communities and the vulnerable, which is sustainable to meet future threats and demands*" (Forensic Capability Network, 2021) by address similar topics, by promoting publicly five objectives:

- Promote sustainability and forward-looking commercial marketplace – in order to effectively forecast and communicate demand.
- Prepare for accreditation (ISO 17025 / ISO 17020) – to meet standards that are in place and assure operational and reputational requirements.
- Endorse cohesion amongst the forensic community – to obtain financial

and operational gains.

- Increase technological engagement – to keep pace with new technology
- Reinforce workforce – in order to be capable to respond in a way that is sustainable.

6.4. Experts' motivation and Need for Cognition

The study conducted by Charlton, Fraser-Mackenzie and Dror (2013) showed that focusing on examiners' motivation has the potential to better understand their work and improve their practices. The studies conducted in Chapters 4 and 5 analysed examiners' work and associated it with their intrinsic motivation and Need for Cognition.

Research has demonstrated that intrinsic motivation and the level of Need for Cognition that individuals show are factors that influence one's performance as well as the possibility to apply heuristics in their reasoning during a decision-making process (Coelho, Hanel & Wolf, 2018; Hoffman, 2015). For those reasons, Chapter 4 observed, in a novel way, the levels of intrinsic motivation and Need for Cognition that fingerprint examiners presented when carrying out the experimental task.

According to previous research, experienced examiners are more accurate than less-experienced examiners (e.g. trainees) (Thompson, Tangen & McCarthy, 2011). Guidelines promote bureaus to follow a numerical approach and to be accredited in order to provide higher quality within their reports. Hence, It could be expected that levels of intrinsic motivation and Need for Cognition could be different amongst participants with different levels of experience, who followed different methodological approaches and were at bureaus that differed in their accreditation. However, in Chapter 4 no significant differences were observed between examiners' motivation and Need for Cognition related to work characteristics such as level of experience, methodological approaches or accreditation of their bureau. This suggests that studies should be focusing on examiners' personality, rather than only on professional competencies.

Chapter 5 demonstrated novel insights related to examiners' motivation and Need for Cognition. During interviews, a majority of participants mentioned that they

were motivated by external sources such as contextual information, being involved in the Criminal Justice System or financial rewards. This observation may suggest that research should start focusing also on extrinsic motivation factors that fingerprint examiners have in order to understand their motivation as a whole. Another observation of interest was the fact that a number of fingerprint examiners that mentioned they were motivated by intrinsic motivational factors (e.g. comparing the task of fingerprint comparisons to a puzzle) and also that they enjoyed that kind of task specifically, suggesting high levels of Need for Cognition. This was observed in previous research that described individuals with high levels of intrinsic motivation to show high levels of Need for Cognition (Gomes, Santos, Gonçalves, Orgambidez-Ramos & Giger, 2013). Hence, based on the findings in this thesis, another suggestion that can be made is to focus on examiners who claim to be motivated by the task itself and learn from them how to motivate others (that have higher levels of extrinsic motivation) to acknowledge the importance of the task of comparing fingerprints and to enable them to enjoy that task in order to increase their Need for Cognition.

6.5. Other considerations of interest

In addition to the considerations discussed in section 6.2. and 6.3., there are two other considerations that this thesis raises which are relevant to the fingerprint community, as well as academia and policymakers; (i) the lack of funding is a fundamental issue facing the forensic community, and (ii) the proximity between the fingerprint community and academia.

It appears that fingerprint examiners are keen to learn new techniques, improve and increase their knowledge, and try out new methodologies. This is in accordance with the research that already stated the need to rethink training for fingerprint examiners related to areas such as influences of contextual information and cognitive bias (Dror, Charlton & Péron, 2006). However, one novel feature is suggested by this thesis; the introduction of having fingerprint examiners actively engaged in the (i) development and (ii) deployment of new training within their bureaus. This suggestion is based on the number of participants that mentioned they felt a lack of training within their practice associated with their interest in

developing new knowledge related to cognitive bias and other topics such as methodological approaches and accreditation.

The need to promote the forensic community to work closely with academia is not new as research has suggested more than once that this should happen more often (Beresford, Stotesbury, Langer, Illes, Kyle & Yamashita, 2019; Mnookin et al., 2011). However, it seems that it is not due to lack of interest or knowledge that this link is not well developed. A fact that corroborates this, is, for instance, the diversity of the sample in Chapter 4 where 67 fingerprint examiners from 15 forensic bureaus based in 9 countries were willing to participate in this research. Additionally, in Chapter 5, it was noted that fingerprint examiners expressed interest in research and wanted to be part of it. However, participants mentioned that there is a lack of opportunity to engage with research or even to carry out their research in-house. Hence, advancements are expected and would be positive, however, as mentioned by participants in Chapter 5, more effort to involve these professionals in research initiatives is also needed.

6.6. Contributions

Although a great effort was made to complete the three studies within this thesis, and an even a greater effort was made by all participants that generously offered their support as participants, there are still gaps that should be highlighted in order to motivate further studies.

Regarding error mitigation techniques, the thesis explored new ways to comprehend how methodologies can be a focus of change. A great amount of previous research has stated that cognitive bias can be promoted by contextual information. Possibly one of the greatest contributions of this thesis was the identification of the influences that different types of contextual information seem to have on a simulation of the Verification phase of the ACE-V process that fingerprint examiners follow in their everyday work.

As mentioned previously, the aim of this thesis was not to ignore the work conducted in the field of cognitive bias within the domain of fingerprint analysis (Executive Office of the President, 2016; Kassin et al., 2013; FSR, 2013;

Langenburg, Champod & Wertheim, 2009; NAS, 2009). On the contrary, this thesis aims to promote it, but enable a more nuanced discussion within it. The studies conducted in this thesis identified that not all contextual information has a negative influence on experts' performance. Another contribution within this matter is related to the sample of this thesis which included fingerprint examiners from various forensic bureaus based in different countries. This enabled a more global set of results and to suggest that findings are not only relevant to a single type of practice but for a wider range of practitioners worldwide.

Another strength of this thesis is the fact that in the quantitative studies, all stimuli that were used was created by the researcher, enabling analysis of accuracy in the comparison trials not by proxy (i.e. when a pool of fingerprint examiners say it is a match or a non-match and the researcher will take the trial as a match) but rather to follow previous research (Mikaelyan & Bigun, 2012; Roussev, 2011; Tear, Thompson & Tangen, 2010) and to create a ground truth database that allowed the researcher to know exactly what participants have assessed in each trial.

Possibly the most important contribution that was made in this thesis was related to capturing the voices of fingerprint examiners. Qualitative research within the field of forensics, where the focus is related to what forensic practitioners think and feel about their work seems to have been largely neglected. This thesis aimed to understand fractured topics such as the impression examiners had of influences related to contextual information and what they thought about other negative issues that need addressing. The novelty of this approach was the ability to combine it with the quantitative results and triangulate the findings to enrich the overall conclusions.

6.7. Limitations and further studies

As in every research, it is important to acknowledge the limitations which can be improved upon in further research.

The use of online experiments, such as the study described in Chapter 3, has been challenged (Murray, Khadjesari, White, Kalaitzaki, Godfrey, McCambridge, Thompson & Wallace, 2009) based on a number of issues such as recruitment,

randomization, fidelity, participation retention and data quality. It was not possible to validate if participants carried the tasks alone or even if they carried out the tasks in a similar context, i.e. without noise or distractions. Possibly this affected participants' responses, even though it was a method that enabled recruitment of a large number of participants. A potential study for the future could be the replication of Chapter 3 using laboratory conditions in order to observe whether laypeople and fingerprint examiners differ as they did in this thesis regarding their accuracy and response time.

The stimuli in Chapter 3 should also be acknowledged in this thesis' limitations. Both artificial fingerprints and excerpts of text were simulations, and therefore will fundamentally differ from the real world. Artificial fingerprints provided the possibility to hold a ground truth dataset, which was also used in research conducted with fingerprint examiners (Mikaelyan & Bigun, 2012) as well as in other fields Roussev (2011). Although the dataset of artificial fingerprints was validated by senior fingerprint experts, this type of stimulus was not retrieved from real cases. Hence, there was a gap between reality and the conditions that participants were presented with during the study conducted in Chapter 3. Similar to artificial fingerprints, using excerpts of text also has limitations due to the fact that this type of stimulus is entirely unlike fingerprint stimuli. Therefore, the use of excerpts of text was a proxy pattern recognition stimuli. In addition to the possibility that the text stimuli lack sufficient ecological validity in this experimental context, it is also potentially problematic that none of the participants in this experience were specifically trained in analysing text-based stimuli. However, this lack of training limitation may also be a benefit, as it levels the playing field between experts and laypeople in the experimental task.

The task in Chapter 4 had limitations regarding the time limit that participants had to complete each trial. When conducting the ACE-V process, fingerprint examiners have more than 30 seconds to assess fingerprints. However, repetition was suggested by previous research (Tangen, Thompson & McCarthy, 2011) and this type of design was a procedure to make participants repeat the same task several times. In this study, the aim was not to understand how accurate fingerprint examiners would be in the first place, but how their accuracy and response time suffered from the presence of different types of contextual information. It is

expected that if given standard time to make comparisons, accuracy would be greater, nevertheless, a further study that could be attempted to conduct is the replication of the study in Chapter 4 using standard time in order to observe if in real conditions different types of contextual information would affect differently accuracy and response time of examiners.

Although the advantages that the use of artificial fingerprints in Chapter 3 and 4 had, there were also limitations regarding its use. Conducting the trials using artificial fingerprints reduced the proximity to reality. However, it brought the possibility to access to a ground truth database where it could be possible to be 100% sure about if the pairs matched or not.

Finally, further qualitative studies engaging fingerprint examiners as well as policymakers, regulators and management teams attempting to answer questions that were raised in this thesis, such as quality standards, methodological approaches, recruitment and training would be a very welcome contribution to the literature. Promoting initiatives like this would increase knowledge in the field, motivate examiners, and address issues raised by recent research (Beresford, Stotesbury, Langer, Illes, Kyle & Yamashita, 2019) that claim the bridge between research and practice should be diminished.

6.8 Conclusions

The studies presented within this thesis have demonstrated that both fingerprint examiners and laypeople are influenced from cognitive bias, suggesting that cognitive bias associated with contextual information influence individuals' decision-making processes regardless of expertise. The influences which were found differed regarding the type of contextual information that was presented to either experts or laypeople in terms of accuracy and response time. This suggests that different types of contextual information have different implications within individuals' decision-making processes regardless of the field of expertise one has.

Aiming to better understand this phenomenon, one study carried out only with fingerprint examiners explored how different contextual information influenced fingerprint examiners' reasoning during a verification task – similar to the

Verification phase of the ACE-V process. In this study, similar results were found, suggesting that different types of contextual information have different influences on examiners' accuracy and response time. Interestingly, contextual information related to the type of crime showed no significant effect when compared to control trials that had no contextual information. The trials where examiners had lower accuracy and higher response times were trials where participants had access to previous conclusions from another examiner. These two observations challenge the setting where fingerprint examiners work as currently, the majority of fingerprint bureaus provide no information related to the type of crime associated to a case and give access to previous conclusions from other fingerprint examiners.

Regarding the type of contextual information that fingerprint examiners have access to, studies in this thesis investigated this in relation to examiners' motivation and level of Need for Cognition. According to examiners' opinions, the majority of participants in this thesis claimed that they were motivated by external sources (e.g. type of crime), challenging the guideline that states bureaus should provide no information related to the case that examiners are working on.

It was acknowledged by this thesis the value of guidelines that require fingerprint bureaus to be accredited and to follow a holistic approach. However, results from the studies within this thesis challenged these assumptions too, showing that there were no significant differences regarding accuracy and response times between examiners that followed different methodological approaches or were working in accredited or non-accredited bureaus. This was possible in this thesis since the samples were from a wide range of fingerprint bureaus internationally. Results indicated that further studies should be carried out and should focus specifically on differences between bureaus that differ in their methodological approach and accreditation level in order to provide new insights to achieve optimal work settings.

Finally, this thesis also demonstrated that examiners are keen to be involved in new advancements within the field of fingerprint comparisons. However, reflections regarding their work conditions need to be in place either by future researchers or by policy decision-makers. One of the suggestions that this thesis provided was to have increased involvement of fingerprint examiners in the discussions related to

their work conditions. If fingerprint examiners are involved, they (i) may have a different perspective and understanding regarding policy-makers' decisions, (ii) provide insights that only individuals who carry out fieldwork can provide, (iii) start to feel empowered within new developments, and (iv) become more dynamic towards activities related to their daily work such as the implementation of proficiency testing, recruitment or research.

REFERENCES

Ackerman, P. L., & Beier, M. E. (2006). Methods for Studying the Structure of Expertise: Psychometric Approaches. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), The Cambridge handbook of expertise and expert performance (p. 147–165). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.009

Adams, W. C. (2015). Conducting Semi-Structured Interviews. In K. E. Newcomer & H. P. Hatry (Eds.), *Handbook of Practical Program Evaluation*. https://doi.org/10.1002/9781119171386.ch19

Alexander, K. L. (2015a, April 27). National accreditation board suspends all DNA testing at D.C. crime lab. *The Washington Post*. Retrieved from http://www.thewashingtonpost.com

Alexander, K. L. (2015b, March 5). Prosecutor criticize D.C. crime lab's handling of some DNA evidence. *The Washington Post*. Retrieved from http://www.thewashingtonpost.com

Alexander, A., Botti, F., Dessimoz, D. & Drygajlo, A. (2004). The effect of mismatched recording conditions on human and automatic speaker recognition in forensic applications. *Forensic Science International*. *2, 146,* 95-99 DOI: 10.1016/j.forsciint.2004.09.078

American Society of Crime Laboratory Directors / Laboratory Accreditation Board (2011). Supplemental Requirements for the Accreditation of Forensic Science Testing Laboratories.

Aquino, A., Haddock, G., Maio, G. R., Wolf, L. J., Alparone, F. R. (2016). The role of affective and cognitive individual differences in social perception. *Personality and Social Psychology Bulletin. 42,* 798-810. doi:10.1177/0146167216643936

Aronson, E. (1999). The Social Animal (8th edition). Worth Publishers: New York

Ask, K., Rebelius, A. & Granhag, P. A. (2008). The 'Elasticity' of Criminal Evidence: A Moderator of Investigator Bias. *Applied Cognitive Psychology*. *Vol. 22,* 1245-1259

Bagby, R. M., Taylor, G. J., Ryan, D. (1986). Toronto Alexithymia Scale: Relationship with personality and psychopathology measures. *Psychotherapy and Psychosomatics*. *45*, 207-215.

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G. & Frith, C. (2012). Together, slowly but surely: The role of social interaction and feedback on the build-up of benefit in collective decision-making. *Journal of Experimental Psychology: Human Perception and Performance*. *Vol 38(1)*, 3-8

Barnes, J. G. (2011). The Fingerprint Sourcebook. U.S.: *National Institute of Justice [NIJ]*.

Baron, J. (2014). Heuristics and biases. In E. Zamir & D. Teichman (Eds.), Oxford handbooks. The Oxford handbook of behavioral economics and the law (p. 3–27). Oxford University Press.

Bateman, I. J. & Mawby, J. (2004). First impressions count: interviewer appearance and information effects in stated preference studies. *Ecological Economics*. *49(1)*, 47-55

BBC. (2011). Inquiry on Shirley McKie case blames 'human error'. Retrieved from https://www.bbc.com/news/uk-scotland-glasgow-west-16181875

Ben-Zur, H. & Breznitz, S. J. (1981). The effects of time pressure on risky choice behavior. *Acta Psicologica*. 47: *89-104*

Beresford, D. V., Stotesbury, T., Langer, S. V., Illes, M., Kyle, C. J. & Yamashita, B. (2020). Bridging the gap between academia and practice: Perspectives from two large-scale and niche research projects in Canada. *Science and Justice*. *60*, 95-98

Biedermann, A., Bozza, S., Taroni, F. & Vuille, J. (2019). Are Inconclusive Decisions in Forensic Science as Deficient as They Are Said to Be? *Frontiers in Psychology*. *19*, doi.org/10.3389/fpsyg.2019.00520

Bless, H., Wänke, M., Bohner, G., Fellhauer, R. F., Schwarz, N. (1994). Need for cognition: Eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben [Presentation and validation of a German version of the Need for Cognition Scale]. Zeitschrift Für Sozialpsychologie, 25, 147-154.

Bloom, B. S. (Ed.) (1985). Developing talent in young people. New York: Ballantine.

Bloom, B. S. (1982, April). The role of gifts and markers in the development of talent. Exceptional Children, 48(6), 510–521.

Bonefeld, M. & Dickhäuser, O. (2018). (Biased) Grading of Students' Performance: Students' Names, Performance Level, and Implicit Attitudes. *Frontiers in Psychology*. doi.org/10.3389/fpsyg.2018.00481

Boud, D. and Molloy, E. (2013). Decision-making for feedback. Boud, D. and Molloy, E. (Eds.). Feedback in Higher and Professional Education, London: Routledge, 202-217

Braun, V. and Clarke, V. (2006) 'Using Thematic Analysis in Psychology' Qualitative Research in Psychology 3(2): 77-101.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, *81*, 211-241

Bryman, A. (2012) *Social Research Methods* (4rd Edition), Oxford: Oxford University Press.

Bulmer, M. (2008) 'The Ethics of Social Research' in N. Gilbert (3rd Eds) *Researching Social Life*, London: Sage, 145-162

Bundeskriminalamt. (2010). Standard des daktyloskopischen Identitätsnachweises. Frankfurt

Busey, T. A. & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research*. *45(4)*, 431-448

Burgman, M. A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., Findler, F., Rumpff, L. & Twardy, C. (2011). Experts Status and Performance. *PLoS ONE*. 6(7): e22998. doi:10.1371/journal.pone.0022998

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin. 119*, 197-253.

Cacioppo, J. T., Petty, R. E., Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*. *48*, 306-307. doi:10.1207/s15327752jpa4803_13

Cacioppo, J. T., Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*. *42*, 116-131. doi:10.1037/0022-3514.42.1.116

Campbell, G. & Waters, D. A. (2013). Making decisions in emergency surgery. *ANZ Journal of Surgery*. *83(6)*: 429-433

Cappelli, R. (2015). SFinGe, In Z. Stan & A. K. Jain (Eds.), Encyclopaedia of Biometrics (second edition). Springer, 2015.

Chai, S. C., Teoh, R. F., Razaob, N. A. & Kadar, M. (2017). Work motivation among occupational therapy graduates in Malaysia. *Hong Kong Journal of Occupational Therapy. 30*, 42-48

Champod, C., Lennard, C., Margot, P. A. & Stoilovic, M. (2016). Fingerprint and other ridge skins impressions. Boca Raton, FL: CRC Press

Champod, C. (2014). Research focused mainly on bias will paralyse forensic science. *Science & Justice*. *54(2),*107-109

Champod, C. (2009). Identification and Individualization. In A. Jamieson & A. Moenssens (Eds.). Wiley Encyclopedia of Forensic Science. https://doi.org/10.1002/9780470061589.fsa134

Champod, C. & Chamberlain, P. (2009). Fingerprints. In J. Fraser & R. Williams (Eds.). *Handbook of Forensic Science*. Devon: Willan Publishing

Charlton, D., Fraser-Mackenzie & Dror, I. (2010). Emotional Experiences and Motivating Factors Associated with Fingerprint Analysis. *Journal of Forensic Sciences. Vol. 55, No. 2,* 385-393

Chi, M. T. H. (2006). Laboratory Methods for Assessing Experts' and Novices' Knowledge. In K. A. Ericsson, N. Charness, P. Feltovich, & R. R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 167-184). Cambridge: Cambridge University Press

Clarke, J. R., Spejewski, B., Gertner, A. S., Webber, B. L., Hayward, C. Z., Santora, T. A., Wagner, D. K., Baker, C. C., Champion, H. R., Fabian, T. C., Lewis Jr., F. R., Moore, E. E., Weigelt, J. A., Eastman, A. B. & Blank-Reid, C. (2000). An Objective Analysis of Process Errors in Trauma Resuscitations. *Academic Emergency Medicine*. *7(11)*, 1303-1310

Cochran, W. G. (1943). The comparison of different scales of measurement for experimental results. *Annals of Mathematical Statistics*. *14*, 205–216

Coelho, G. L. H., Hanel, P. H. P. & Wolf, L. J. (2018). The Very Efficient Assessment of Need for Cognition: Developing a Six-Item Version. Assessment. https://doi.org/10.1177/1073191118793208

Cohen, L., Manion, L. & Morrison, K. (2007). Research Methods in Education. London: Routledge

Cohen, M. S. (1993). The naturalistic basis of decision biases. In G. A. Klein, J. Orasanu, R. Calderwood,& C. E. Zsambok (Eds.), Decision making in action: Models and methods (pp. 51-99). Norwood, NJ:Ablex.

Cohen, A. R., Stotland, E., & Wolfe, D. M. (1955). An experimental investigation of need for cognition. *The Journal of Abnormal and Social Psychology*. *51(2)*, 291–294. https://doi.org/10.1037/h0042761

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*, 117-150

Cole, S. A. (2005). More than Zero: Accounting for Error in Latent Fingerprint Identification. *Journal of Criminal Law and Criminology*. *95(3)*, 985-1078

Cole, S. A. (2002). Suspects Identities: A History of Fingerprints and Criminal Identification. London: Harvard University Press

Cooper, G. S., & Meterko, V. (2019). Cognitive bias research in forensic science: A systematic review. Forensic Science International, 297, 35–46. https://doi.org/10.1016/j.forsciint.2019.01.016

Cooper, D. (2002). Work motivation: personal and situational influences on behaviour. In T. Hannagan, *Management, Concepts and Practices* (pp. 313-348). Pearson Education Limited: Edinburgh

Cormier, E. M., Pickett-Hauber, R., & Whyte, I. V. J. (2010). Cognitions and Clinical Performance: A Comparison of High and Low Performing Baccalaureate Nursing Students. *International Journal of Nursing Education Scholarship*. 7(1), 1-18

Correia, P. & Pinheiro, M. F. (2013). Perspectivas Actuais da Lofoscopia: Aplicação Criminal e Civil do Estudo de Impressões Epidérmicas. In M. F. Pinheiro (Eds.) *Ciências Forenses ao Serviço da Justiça*. Lisboa: Pactor

Creswell, J. W. (2014). Research Design: Qualitative, Quantitative and Mixed Methods Approaches (4th ed.). Thousand Oaks, CA: Sage

Creswell, J.W. & Plano Clark V.L., (2011) Designing and Conducting Mixed Methods Research. London: Sage.

Creswell, J. W., Klassen, A. C., Plano Clark, V. L. & Smith, K. C. (2010). Best Practices for Mixed Methods Research in the Health Sciences. Office of Behavioral and Social Sciences Research (OBSSR)

Creswell, J. W. (2007). Qualitative inquiry and research design: Choosing among five approaches (2nd ed.). Sage Publications, Inc.

Croskerry P, Singhal, G. & Mamede, S. (2013). Cognitive debiasing: origins of bias and theory of debiasing. *BMJ Quality and Safety*. doi:0:1–7. doi:10.1136/bmjqs-2012-001712 1

Culhane, S. E., Morera, O. F., Hosch, H. M. (2004). The factor structure of the need for cognition short form in a Hispanic sample. Journal of Psychology, 138, 77-90. doi:10.3200/JRLP.138.1.77-90

Cummins, H. (1935). Attempts to Alter and Obliterate Finger-Prints. *Journal of the American Institute of Criminal Law and Criminology*. *Vol. XXV*, 982-991

Darley, J. M. & Gross, P. H. (1983). A Hypothesis-Confirming Bias in Labeling Effects. *Journal of Personality and Social Psychology*. *44, 1,* 20-33

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*. *11*, 227–268.

Department of Justice & National Institute of Standards and Technology. (2016). Proficiency Testing in Forensic Science. National Commission on Forensic Science. Retrieved from https://www.justice.gov/archives/ncfs/page/file/831806/download

DiCicco-Bloom, B. & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education. 40, 4,* 314-321

Dror, I. & Langenburg, G. (2019). “Cannot Decide”: The Fine Line Between Appropriate Inconclusive Determinations Versus Unjustifiably Deciding Not To Decide. *Journal of Forensic Sciences*. *64(1),* https://doi.org/10.1111/1556-4029.13854

Dror, I. (2013). Practical Solutions to Cognitive and Human Factor Challenges in Forensic Science. *Forensic Science Policy & Management. 4(3-4),* 1-9

Dror. I., Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science and Justice*. *Vol. 52, 4,* 204-208

Dror, I. (2012). Letter to the Editor—Combating Bias: The Next Step in Fighting Cognitive and Psychological Contamination. *Journal of Forensic Sciences*. https://doi.org/10.1111/j.1556-4029.2011.01940.x

Dror, I., Wertheim, K., Fraser-Mackenzie, P. & Walajtys, J. (2012). The Impact of Human-Technology Cooperation and Distributed Cognition in Forensic Science: Biasing Effects of AFIS Contextual Information on Human Experts. *Journal of Forensic Sciences*. *Vol. 57, No. 2,* 343-352

Dror, I. Cole, S. (2010). The vision in "blind" justice: Expert perception, judgement, and visual cognition in forensic pattern recognition. *Psychonomic Bulletin & Review*. *Vol. 17, 2,* 161-167

Dror, I., Charlton, D. & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*. *156, 1,* 74-78

Dyer, A. G., Found, B. & Rogers, D. (2006). Visual Attention and Expertise for Forensic Signature Analysis. *Journal of Forensic Sciences*. *51, 6,* 1397-1404

Earwaker, H. Nakhaeizadeh, S., Smith, N. M. & Morgan, R. M. (2020). A cultural change to enable improved decision-making in forensic science: A six phase approach. *Science & Justice*. *60, 1,* 9-19

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 & Justice*. *55, 4,* 239-247

Earwaker, H., Charlton, D. & Bleay, S. (2015). Fingerprinting – the UK Landscape: Processes, Stakeholders and Interactions. Forensic Science Special Interest Group, Knowledge Transfer Network

Egglin, T. K. & Feinstein, A. R. (1996). Context bias. A problem in diagnostic radiology. *JAMA. 4;276(21)*: 1752-1755

Einhorn, J. (1974). Expert judgment: Some necessary conditions and an example. *Journal of Applied Psychology. 59,* 562–571

Elmir, R., Schmied, V., Jackson, D. & Wilkes, L. (2011). Interviewing people about potentially sensitive topics. Nurse Researcher. *19, 1,* 12-16

Enough, B., & Mussweiler, T. (2001). Sentencing Under Uncertainty: Anchoring Effects in the Courtroom1. Journal of Applied Social Psychology, 31(7), 1535–1551. https://doi.org/10.1111/j.1559-1816.2001.tb02687.x

Ericsson, K. A. (2006). The Influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. Feltovich, & R. R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 287-301). Cambridge: Cambridge University Press

Ericsson, K. A., Charness, N., Feltovich, P. & Hoffman, R. R. (2006). Cambridge Handbook of Expertise and Expert Performance. Cambridge: Cambridge University Press

Ericsson, K. A., & Smith, J. (1991). Prospects and limits in theempirical study of expertise: An introduction. In K. A. Ericsson& J. Smith (Eds.), Toward a general theory of expertise:Prospects and limits(pp. 1–38). Cambridge, UK: Cambridge University Press

European Network of Forensic Science Institutes (2015). Best Practice Manual for Fingerprint Examination. Retrieved from ENFSI at http://enfsi.eu/wp-content/uploads/2016/09/6._fingerprint_examination_0.pdf

Ettenson, R., Shanteau, J., & Krogstad, J. (1987). Expert judgment: Is more information better? *Psychological Reports. 60*, 227–238

Evett, I. & Williams, R. L. (1996). A Review of the Sixteen Point Fingerprint Standard in England and Wales. Journal of Forensic Identification. *46(1)*, 49-73

Executive Office of the President. (2016). REPORT TO THE PRESIDENT
Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods.

FBI National Press Office. (2004). Statement on Brandon Mayfield Case. Retrieved from https://archives.fbi.gov/archives/news/pressrel/press-releases/statement-on-brandon-mayfield-case

Field, A. (2009). DISCOVERING STATISTICS USING SPSS (and sex and drugs and rock 'n' roll). Thousand Oaks, CA: SAGE

Forensic Capability Network | Why FCN?. (2020). Forensic Capability Network. https://www.fcn.police.uk/why-fcn

Forensic Science Rregulator. (2020). Cognitive Bias Effects Relevant to Forensic Examinations FSR-G-217. UK: FSR, Home Office

Forensic Science Regulator. (2019). Forensic science and the criminal justice system: a blueprint for change - Forensic Science Regulator response. UK: FSR, Home Office

Forensic Science Regulator. (2018). Forensic Science Regulator annual report 2017. UK: FSR, Home Office

Forensic Science Regulator. (2017). Codes of Practice and Conduct for forensic science providers and practitioners in the Criminal Justice System, Issue 4. UK: FSR, Home Office

Forensic Science Regulator. (2015). Cognitive Bias Effects Relevant to Forensic Science Examinations FSR-G-217 Issue 1.UK: FSR, Home Office

Forensic Science Regulator. (2013). Codes of Practice and Conduct – Draft guidance and standards on the analysis and comparison of fingerprints. UK: FSR, Home Office

Forensic Science Regulator. (2011). Developing a Quality Standard for Fingerprint Examination. *Standards for Fingerprint Identification*. UK: FSR, Home Office

Forsterlee, R., Ho, R. (1999). An examination of the short form of the Need for Cognition Scale applied in an Australian sample. Educational and Psychological Measurement, 59, 471-480. doi:10.1177/00131649921969983

Fylan, F. (2006). Semi-structured interviewing. In J. Miles & P. Gilbert (Eds.), A handbook of research methods for clinical and health psychology (pp. 65-77). Oxford, UK: Oxford University Press.

Gagné, M. & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*. *26(4),* 331-362
Galton, F. (1892). Finger Prints. London: MacMillan And Co.

Gallop, A. (2020, February 13). If the UK cares about justice, it must fund forensic services properly | Angela Gallop. Retrieved December 20, 2020, from http://www.theguardian.com/commentisfree/2020/feb/13/uk-justice-fund-forensic-services-budget-cuts

Garret, B. L. & Neufeld, P. J. (2009). Invalid Forensic Science Testimony and Wrongful Convictions. *Virginia Law Review*. *Vol. 95, No. 1,* 1-97

Gavett, G. (2012, April 12). Can Unconscious Bias Undermine Fingerprint Analysis? Retrieved December 20, 2020, from http://www.pbs.org/wgbh/frontline/article/can-unconscious-bias-undermine-fingerprint-analysis/

Gibb, F. (2019). 'Fund forensics or more crimes will go unsolved'. *The Times*. Retrieved from https://www.thetimes.co.uk/article/fund-forensics-or-more-crimes-will-go-unsolved-x50npzd5c

Gilbey, A. & Hill, S. (2012). Confirmation Bias in General Aviation Lost Procedures. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.2860

Gilovich, T., & Savitsky, K. (2002). Like goes with like: The role of representativeness in erroneous and pseudo-scientific beliefs. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (p. 617–624). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.036

Ghasemi, A. & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology & Metabolism*. 10(2): *486-489,* doi: 10.5812/ijem.3505

Golde, R.A. (1970). Can you be sure of your experts? New York: Award Books.

Gomes, A., Santos, J. V., Gonçalves, G., Orgambidez-Ramos, A. & Giger, J.-C. (2013). Estudo de validação da Escala de Necessidade de Cognição com amostra portuguesa. *Avaliação Psicológica*. *12(2),* 179-192

Gonçalves, F. V. (2017). Differences between fingerprint agencies across numerical and holistic approaches. *Fingerprint Whorld, Chartered Society of Forensic Sciences*. 2, 18-21

Gonzalez, C., Fakhari, P., & Busemeyer, J. (2017). Dynamic Decision Making: Learning Processes and New Research Directions. *Human Factors*. 59 (5), *713–721*. doi.org/10.1177/0018720817710347

Gosling, S. D. & Mason, W. (2015). Internet Research in Psychology. *Annual Review of Psychology. 66:1*, 877-902

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. American Psychologist, 59(2), 93-104.

Gouveia, V. V., Mendes, L. A. de C., Soares, A. K. S., Monteiro, R. P., Santos, L. C. de O. (2015). Need for Cognition Scale (NCS-18): Effect of Negative Items in its Factorial Structure. Psicologia: Reflexão E Crítica, 28, 425-433. doi:10.1590/1678-7153.201528301

Greene, J. C., Caracelli, V. J. & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*. *Vol. 11, No. 3*, 255-274

Gruber, H. E. (1982). On the hypothesized relation between giftedness and creativity. In D. H. Feldman (Ed.), Developmental approaches to giftedness and creativity (pp. *7–29*). San Francisco: Jossey-Bass.

Haddock, G., Maio, G. R., Arnold, K., Huskinson, T. (2008). Should persuasion be affective or cognitive? The moderating effects of need for affect and need for cognition. *Personality and Social Psychology Bulletin*. *34*, 769-778.

Hall, L. J. & Player, E. (2008). Will the introduction of an emotional context affect fingerprint analysis and decision-making? *Forensic Science International*. *181(1-3)*, 36-39

Harrel, M. C. & Bradley, M. A. (2009). Data Collection Methods: Semi-Structured Interviews and Focus Groups. NATIONAL DEFENSE RESEARCH INSTITUTE. Retrieved from https://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR718.pdf

Healy, A. F., Kole, J. A. & Bourne, L. E. (2014). Training Principles to Advance Expertise. *PubMed*. DOI: 10.3389/fpsyg.2014.00131

Hewson, Claire, M. Peter Yule, Dianna Laurent & Carl M. Vogel (eds). 2003. Internet research methods: A practical guide for the social and behavioural sciences. New technologies for social research. London, UK: Sage

Hoffman, B. (2015). Motivation for Learning and Performance. New York: Academic Press

Hoffman, R. R. (1998). How Can Expertise be Defined? Implications of Research From Cognitive Psychology. In R. Williams, W. Faulkner & J. Fleck (Eds.), (1998). Exploring Expertise (pp. 81–100). Edinburgh, Scotland: University of Edinburgh Press.

Holland, J. L. (1997). Making vocational choices: A theory of vocational personalities and work environments. Psychological Assessment Resources.

Huber, O. & Kunz, U. (2007). Time pressure in risky decision-making: effects on risk defusing. *Psychological Science*. *49*: 415-426

IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

Jamshed, S. (2014). Qualitative research method-interviewing and observation. *PubMed*. DOI: 10.4103/0976-0105.141942

Jessup, R. K., Bishara, A. J. & Busemeyer, J. R. (2008). Feedback Produces Divergence From Prospect Theory in Descriptive Choice. *Psychological Science*. *Vol 19(10)*, 1015-1022

Johnson, R.B. and Christensen, L.B. (2008) Educational Research: Quantitative, Qualitative, and Mixed Approaches. 3rd Edition, Sage Publications, Inc., Lose Angeles.

Jones, G. R. & George, J. M. (2008). Motivation and Performance. In G. R. Jones & J. M. George (Eds.), Contemporary Management. McGraw Hill: New York

Juric, M. (2016). The role of the need for cognition in the university students' reading behaviour. *Proceedings of ISIC: the information behaviour conference, Zadar, Croatia, 20-23 September, 2016: Part 2.*

Kahneman, D. (2011a). Expert Intuition: When Can We Trust It?. In D. Kahneman, *Thinking Fast and Slow*. (pp.234-245). New York: Farrar, Straus and Giroux

Kahneman, D. (2011b). The illusion of validity. In D. Kahneman, *Thinking Fast and Slow*. (pp.234-245). New York: Farrar, Straus and Giroux

Kahneman, D. (1991). Judgment and decision making: A personal view. Psychological Science, *2(3),* 142-145.

Kahneman, D., Slovic, P., & Tversky, A. (Ed.). (1982). Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*. *80, 4*, 237-251

Kaplan, J., Ling, S. & Cuellar, M. (2020). Public Beliefs About the Accuracy and Importance of Forensic Evidence in the United States. *Science & Justice*. doi.org/10.1016/j.scijus.2020.01.001

Kaplan, M. F., Wanshula, L. T. & Zanna, M. P. (1993). Time pressure and information integration in social judgment. In O. Svenson & A. J. Maule (Eds). *Time pressure and Stress in Human Judgement and Decision Making*. New York: Plenum

Kao, C. (1994). The concept and measurement of need for cognition: The concept and measurement of need for cognition. Chinese Journal of Psychology, 36, 1-20.

Kardes, F. R., Muthukrishnan, A. V. & Pashkevich, V. (2005). On the Conditions Under Which Experience and Motivation Accentuate Bias in Intuitive Judgement. In

T. Betsch & S. Haberstroh (Eds). *The Routines of Decision Making*. New York: Psychology Press

Kassin, S. M., Dror, I. & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives and proposed solutions. *Journal of Applied Research in Memory and Cognition. Vol. 2,* 42-52

Kanfer, R. (1990). Motivation theory and industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (p. 75–170). Consulting Psychologists Press.

Kaye, D. H. (2009). Probability, Individualization, and Uniqueness in Forensic Science Evidence: Listening to the Academies. Brooklyn Law Review, Vol. 75, No. 4, pp. 1163-1185.

Kearney, E., Gebert, D., Voelpel, S. C. (2009). When and how diversity benefits teams: The importance of team members' need for cognition. *Academy of Management Journal*. *52*, 581-598. doi:10.5465/AMJ.2009.41331431

Kershaw, S. & Lichtblau, E. (2004). "Bomb Case Against Lawyer Is Rejected". *The New York Times*. Retrieved from https://www.nytimes.com/2004/05/25/us/bomb-case-against-lawyer-is-rejected.html

Kerstholt, J. H. & Raaijmakers, G. W. (1997). Decision making in dynamic task environments. In R. Ranyard, R. Crozier & O. Svenson (Eds). *Decision Making: Cognitive models and explanations*. Routledge: London and New York

Klayman, J. (1995). Varieties of Confirmation Bias. *Psychology of Learning and Motivation*. *32*, 385-418

Kleinmuntz, D. N. (1985). Cognitive Heuristics and Feedback in a Dynamic Decision Environment. *Management Science, vol. 31(6),* 680-702

Komarinski, P. (2005). Automated Fingerprint Identification Systems. London: Elsevier Academic Press

Langenburg, G., Champod, C. & Wertheim, P. (2009). Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology when Conducting Fingerprint Comparisons. *Journal of Forensic Science*. *54(3)*, 571-582

Liedtka, J. (2015). Perspective: Linking Design Thinking with Innovation Outcomes through Cognitive Bias Reduction. *Journal of Product Development and Management Association*. *32(6)*, 925-938

Liu, Y.-I., Eveland, W. P. (2005). Education, need for cognition, and campaign interest as moderators of news effects on political knowledge: An analysis of the knowledge gap. *Journalism & Mass Communication Quarterly*. *82*, 910-929. doi:10.1177/107769900508200410

Lodge, C. (2005). From hearing voices to engaging in dialogue: problematising student participation in school improvement. Journal of Educational Change, 6(2), 125–146. doi:10.1007/s10833-005-1299-3

Lundberg, J., Forsell, C., Johansson, J. & Josefsson, B. (2014). The Use of Conflict Detection Tools in Air Traffic Management – an Unobtrusive Eye Tracking Field Experiment During Controller Competence Assurance. Conference: HCI-Aero 2014 - International Conference on Human-Computer Interaction in Aerospace, At Silicon Valley, California, USA

MacCoun, R. & Perlmutter, S. (2015). Hide results to seek the truth. *Nature*. 526(7572), 187-189

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. Behavior Research Methods, *44(2)*, 314-324. doi:10.3758/s13428-011-0168-7

Mason, J. (1994). Linking qualitative and quantitative data analysis. In Analyzing qualitative data, A. Bryman & R. G. Burgess (Eds.) (pp. 89–110). New York: Routledge.

Mantel, S. P. & Kardes, F. R. (1999). The Role of Direction of Comparison Attribute-Based Processing, and Attitude-Based in Consumer Preference. *Journal of Consumer Research. Vol 25,* 335-352

Maule, A. J. & Edland, A. C. (1997). The effects of time pressure on human judgement and decision making. In R. Ranyard, R. Crozier & O. Svenson (Eds). *Decision Making: Cognitive models and explanations*. Routledge: London and New York

McRobert, A. P., Causer, J., Vassiladis, J., Watterson, L., Kwan, J. & Williams, M. A. (2013). Contextual information influences diagnosis accuracy and decision making in simulated emergency medicine emergencies. *BMJ Quality and Safety*. *22(6)*: 478-484

Meuwly, D. (2011). Position of the European Fingerprint Working Group (EFPWG) of the European Network of Forensic Science Institutes (ENFSI) Regarding the NRC report. *Journal Forensic Identification. 61,* 677–679

Mikaelyan, A. & Bigun, J. (2012). Ground truth and evaluation for latent fingerprint matching. *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, 2012, pp. 83-88. doi: 10.1109/CVPRW.2012.6239220*

Mnookin, J. L., Cole, S. A., Dror, E. Fisher, B., Houck, M. M., Inman, K., Kaye, D. H. Koehler, J. J., Langenburg, G., Risinger, D. M., Rudin, N., Siegel, J. & Stoney, D. A. (2011). The Need For a Forensic Culture in The Forensic Sciences. *UCLA Law Review. 58,* 726-779

Mnookin, J. L. (2010). Fingerprint Evidence in an Age of DNA Profiling. *Brooklyn Law Review*. 67

Moertl, P., Bonaceto, C., Estes, S. & Burns, K. (2008). Macrocognition in Systems Engineering: Supporting Changes in the Air Traffic Control Tower. In J. M. Schraagen, L. G. Militello, T. Ormerod & R. Lipshitz (Eds.) *Naturalistic Decision Making and Macrocognition*. Burlington: Ashgate Publishing Company

Montibeller, G. & Winterfeldt, D. (2015). Cognitive and Motivational Biases in Decision and Risk Analysis. *Risk Analysis*. Vol. *35, 7*, 1230-2151

Munro, G. D. & Stansbury, J. A. (2009). The Dark Side of Self-Affirmation: Confirmation Bias and Illusory Correlation in Response to Threatening Information. *Personality and Social Psychology Bulletin*. *35(9),*1143-53

Murray, E., Khadjesari, Z., White, I., Kalaitzaki, E., Godfrey, C., McCambridge, J., Thompson, S. & Wallace, P. (2009). Methodological Challenges in Online Trials. *Journal of Medical Internet Research*. *11(2),* doi:10.2196/jmir.1052

Nagy, G. & Zou, J. (2002). "Interactive visual pattern recognition," Object recognition supported by user interaction for service robots, Quebec City, Quebec, Canada. DOI: 10.1109/ICPR.2002.1048342

National Academy of Sciences [NAS]. (2009). Strengthening Forensic Science In The United States: A Path Forward. *National Research Council of The National Academies.* National Academy Press: Washington, D.C.

National Institute of Standards and Technology. (2018). OSAC Releases an Online Lexicon for the Forensic Sciences. Retrieved from https://www.nist.gov/news-events/news/2018/03/osac-releases-online-lexicon-forensic-sciences

National Institute of Standards and Technology [NIST]. (2012). Expert Working Group on Human Factors in Latent Print Analysis. *Latent Print Examination and Human Factors: Improving the Practice through a System Approach.* U.S. Department of Commerce

National Institute of Standards and Technology [NIST]. (n.d.). OSAC Research and Development Needs. Retrieved from https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/research-and-development-needs

Neumann, C. (2012). Statistics and Probabilities as a Means to Support Fingerprint Examination. In R. S. Ramotowski (Ed). Lee and Gaensslen's Advances in Fingerprint Technology (420-461). CRC Press: Boca Raton

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*. doi.org/10.1037/1089-2680.2.2.175

Nordhall, O. & Knez, I. (2018). Motivation and Justice at Work: The Role of Emotion and Cognition Components of Personal and Collective Work Identity. *Frontiers in Psychology*. *15*, doi.org/10.3389/fpsyg.2017.02307

Office of the Inspector General. (2006). A Review of the FBI's Handling of the Brandon Mayfield Case. Retrieved from https://oig.justice.gov/special/s0601/exec.pdf

O'Neill, E. (2011). Fingerprint evidence 'based on opinion rather than fact'. *The Guardian*. Retrieved from https://www.theguardian.com/uk/2011/dec/14/fingerprint-evidence-opinion-fact

Osborne, N. K., Woods, S., Kieser, J. & Zajac, R. (2014). Does contextual information bias bitemark comparisons? *Science & Justice*. *54(4)*, 267-273

Osman, M. (2012). The Role of Reward in Dynamic Decision Making. *Frontiers in Neuroscience. Vol 6*: 35, doi: 10.3389/fnins.2012.00035

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.). *Cognitive Illusions. A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. Hove and N.Y.: Psychology Press

Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993). The Adaptive Decision Maker. Cambridge: Cambridge University Press

Peachey, P. (2015, January 21). Privatisation of forensic services 'threat to justice' and putting the work in police hands would be 'disastrous', warn experts. *The Independent*. Retrieved from http://www.independent.co.uk

Pervouchine, V. & Leedham, G. (2007). Extraction and analysis of forensic document examiner features used for writer identification. *Pattern Recognition*. 40

Peterson, J.L., Lin, G., Ho, M., Chen, Y., & Gaensslen, R. E. (2003). “The feasibility of external blind DNA proficiency testing. II. Experience with actual blind tests.” *Journal of Forensic Science*. *Vol. 48, No. 1,* 32-40

Pi, D., Parisi, F., & Luppi, B. (2014). Biasing, debiasing, and the law. In E. Zamir & D. Teichman (Eds.), Oxford handbooks. The Oxford handbook of behavioral economics and the law (p. 143–166). Oxford University Press.

Pieters, R. G., Verplanken, B., Modde, J. M. (1987). “Neiging tot nadenken”: Samenhang met beredeneerd gedrag [Leaning toward reflection: The connection with reasoned behavior]. Nederlands Tijdschrift Voor de Psychologie En Haar Grensgebieden, 42(2), 62-70.

Pines, J. M. (2006). Profiles in patient safety: confirmation bias in emergency medicine. *Academic Emergency Medicine*. *13(1)*, 90-94

Proença, T. & Cristina, S. (2013). Motivation in Paid Work for Non-Prot Organisations: The Case of Private Social Solidarity Institutions. *FEP Working Papers. 515*, 1-27

Proctor, R. W. & Vu, K.-P. L. (2006). Stimulus–response compatibility principles: Data, theory, and application Boca Raton, FL: CRC Press.

Pyrek, K. M. (2007). Forensic Science Under Siege: The Challenges of Forensic Laboratories and the Medico-Legal Investigation System. London: Elsevier

Qualtrics software, Version 2017-2018 of Qualtrics. Copyright © 2019 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA. https://www.qualtrics.com

Reips, UD. (2000). The Web Experiment Method: Advantages, Disadvantages, and Solutions. In M. H. Birnbaum (Ed.), Psychology Experiments on the Internet (pp. 89-117). San Diego, CA: Academic Press

Ribeiro, G., Tangen, J. M. & McKimmie, B. M. (2019). Beliefs about error rates and human judgment in forensic science. *Forensic science international*. *297*, 138-147

Roodman, A. (2003). Blind Analysis in Particle Physics. Retrieved from https://www.slac.stanford.edu/econf/C030908/papers/TUIT001.pdf

Roussev, V. (2011). An evaluation of forensic similarity hashes. *Digital Investigation*. 8, S34-S41. DOI: https://doi.org/10.1016/j.diin.2011.05.005

Rovira, E., Cross, A., Leitch, E. & Bonaceto, C. (2013). Displaying Contextual Information Reduces the Costs of Imperfection Decision Automation in Rapid Retasking of ISR Assets. *Human Factors. Vol. 56, No.6,* 1036-1049

Sadowski, C. J., Cogburn, H. E. (1997). Need for cognition in the big-five factor structure. *Journal of Psychology*. *131*, 307-312. doi:10.1080/00223989709603517

Sahakian, B. J. & Labuzetta, J. N. (2013). Bad moves: how decision making goes wrong, and the ethics of smart drugs. London: Oxford University Press.

Saks, M. J. & Koehler, J. J. (2008). The Individualization Fallacy in Forensic Science Evidence. *Vanderbilt law review*. *5(25),* 199-219

Saks, M. J. (2005). The Coming Paradigm Shift in Forensic Identification Science. Science, 309(5736), 892–895.doi:10.1126/science.1111565

Saposnik, G., Redelmeier, D., Ruff, C. C. & Tobler, P. N. (2016). Cognitive biases associated with medical decisions: a systematic review. *Medical Informatics Decision Making*. *3; 16(1): 138*

Seal, A. (2016) 'Thematic Analysis' in N. Gilbert and P. Stoneman (4th Eds) Researching Social Life, London: Sage, 443-460.

Shadish, W., Cook, T., and Campbell, D. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference Boston: Houghton Mifflin.

Shanteau, J. (2015). Why task domains (still) matter for understanding expertise. *Journal of Applied Research in Memory and Cognition*. 4(2015): *169-175*

Shanteau, J. (1992a). Competence in experts: The role of task characteristics. Organizational Behavior and Human Decision Processes, *53*, 252- 266.

Shanteau, J. (1992b). The Psychology of Experts: An Alternative View. In, G. Wright & F. Bolger (Eds.), *Expertise and Decision Support*. *(pp. 11-23)*. New York: Plenum Press

Shanteau, J. (1988). Psychological characteristics and strategies of expert decision makers. *Acta Psychologica*. *68*, 203-215

Shanteau, J. & G. J. Gaeth. (1983). Training expert decision makers to ignore irrelevant information. Alexandria, VA: Army Research Institute.

Shariatmadari, D. (2015). Daniel Kahneman: 'What would I eliminate if I had a magic wand? Overconfidence'. *The Guardian*. Retrieved from https://www.theguardian.com/books/2015/jul/18/daniel-kahneman-books-interview

Shu, M. C. (2015). The Impact of Intrinsic Motivation on The Effectiveness of Leadership Style towards on Work Engagement. *Contemporary Management Research*. *11(4)*, 327-350

Siegrist, M., Keller, C., Kastenholz, H., Frey, S. & Wiek, A. (2007). Laypeople's and Experts' Perception of Nanotechnology Hazards. *Risk Analysis*. *27(1)*, 59-69

Silva, J., Garcia-Marques, T. (2006). Medindo necessidade de cognição e fé na intuição: Tradução e adaptação das duas escalas à população portuguesa [Measuring need for cognition and faith in intuition: translation and adaptation of the two scales to the Portuguese population]. Laboratório de Psicologia, 4, 29-43. doi:10.14417/lp.761

Skånér, Y., Strender, L., & Bring, J. (1998). How do GPs use clinical information in the judgements of heart failure? *Scandinavian Journal of Primary Health Care*. *16*, 95–100.

Smith, J. (2021, January 22). Will Bringing Science Back to the White House Reinvigorate Forensics Reform? Retrieved December 20, 2020, from https://www.americanpeacenik.com/will-bringing-science-back-to-the-white-house-reinvigorate-forensics-reform/

Smith, J. (2019). Fingerprint analysis is high-stakes work — but it doesn't take much to qualify as an expert. *The Intercept*. Retrieved from https://theintercept.com/2019/11/29/fingerprint-examination-proficiency-test-forensic-science/

Smith, L. L. & Bond, J. (2015). Criminal Justice and Forensic Sciences. Palgrave: London

Soubelet, A., Salthouse, T. A. (2016). Does need for cognition have the same meaning at different ages? Assessment, 24, 987-998. doi:10.1177/1073191116636449

Sosniak, L. (2006). Retrospective interviews in the study of expertise and expert performance. In K. A. Ericsson, N. Charness, P. Feltovich, & R. R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 287-301). Cambridge: Cambridge University Press

Sperber, A. D. (2004). Translation and validation of study instruments for cross-cultural research. *Gastroenterology*. *126(1),* 124-128, DOI: https://doi.org/10.1053/j.gastro.2003.10.016

Sterman, J. (1989). Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment. *Management Science*. *Vol 35(3),* 321-339

Stewart, T. R., Roebber, P. J., & Bosart, L. F. (1997). The importance of the task in analyzing expert judgment. *Organizational Behavior and Human Decision Processes.* 69, *205–219.*

Supply Chain Academy. (n.d.). "About the Beer Game." Retrieved June 7, 2019, from https://www.supplychain-academy.net/beer-game/

SWGFAST (2011). Individualization / identification Position statement. Retrieved from http://clpex.com/swgfast/CommentsPositions/120306_Individualization-Identification.pdf

SWGFAST (2002). Friction Ridge Examination Methodology for Latent Print Examiners. Retrieved from http://onin.com/fp/lp_exam_methodology.pdf

Tashakkori, A., & Teddlie, C. (1998). Applied social research methods series, Vol. 46. Mixed methodology: Combining qualitative and quantitative approaches. Sage Publications, Inc.

Tear, M. J., Thompson, M. B. & Tangen, J. M. (2010) The importance of ground truth: an open source biometric repository. In: *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society*, 27 September –1 October, San Francisco, CA, USA.

The Fingerprint Inquiry. (2011). The Fingerprint Inquiry Report. Retrieved from www.thefingerprintinquiryscotland.org.uk

The Innocence Project. (2019, December 14). The cases. https://www.innocenceproject.org/

The Innocence Network. (2019, December 14). Publications: 2019. http://www.innocencenetwork.org.uk/

The National Registry of Exonerations. Exonerations database. Accessed at http://www.law.umich.edu/special/exoneration/Pages/about.aspx

Thompson, W. C. & Newman, E. J. (2015). Lay understanding of forensic statistics: Evaluation of random match probabilities, likelihood ratios, and verbal equivalents. Law and Human Behavior, 39(4): 332-49. doi: 10.1037/lhb0000134

Thompson, M. B., Tangen, J. M. Searston, R. A. (2014). Understanding expertise and non-analytic cognition in fingerprint discriminations made by humans. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2014.00737

Thompson, M. B., Tangen, J. M. & McCarthy, D. J. (2014). Human Matching Performance of Genuine Crime Scene Latent Fingerprints. *Law and Human Behavior*. *Vol. 38, No. 1,* 84-93

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

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

Thompson, W. C., & Cole, S. A. (2005). Lessons from the Brandon Mayfield case. *The Champion*. *29(3)*, 42-4

Todorović, D. (2010). Context effects in visual perception and their explanations. *Review of Psychology*, *17(1),* 17–32.

Torngren, G. & Montgomery, H. (2010). Worse Than Chance? Performance and Confidence Among Professionals and Laypeople in the Stock Market. *Journal of Behavioral Finance*. *5, https://doi.org/10.1207/s15427579jpfm0503_3*

Tremblay, M. A., Blanchard, C. M., Taylor, S., Pelletier, L. G., & Villeneuve, M. (2009). Work Extrinsic and Intrinsic Motivation Scale: Its value for organizational psychology research". *Canadian Journal of Behavioural Science*. *42(1), 70.* doi.org/10.1037/a0018176

Triplett, M. (2012). Is ACE-V a Process or a Method? *IDentification News*. *Vol 42, No. 3,* 6-7

Tsai, K. H., Moskowitz, A. L., Brown, T. E., Park, A. L. & Chorpita, B. F. (2015). Interpreting Progress Feedback to Guide Clinical Decision-Making in Children's Mental Health Services. Administration and Policy in Mental Health and Mental Health Services Research. *Vol 42*(4), 1-8

Tversky, A. & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science, New Series. Vol 211, No. 4481*. 453-458

Tversky, A. & Kahneman, D. (1974). Judgement under and Uncertainty: Heuristics and Biases. *Science, New Series. Vol. 185, No. 4157*, 1124-1131

Tversky, A., & Kahneman, D. (1971). The belief in the 'law of small numbers.'. *Psychological Bulletin*. 76, 105-110.

van Rossum, G. (1995). Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam.

Veletsianos, G. (2010). Contextually relevant pedagogical agents: Visual appearance, stereotypes, and first impressions and their impact on learning. *Computers & Education*. 55(2), 576-585.

Verplanken, B. & Svenson, O. (1997). Conceptualisations and effects on decision processes. In R. Ranyard, R. Crozier & O. Svenson (Eds). *Decision Making: Cognitive models and explanations*. Routledge: London and New York

Viera, A. J. & Bangdiwala, S. I. (2007). Eliminating bias in randomized controlled trials: importance of allocation concealment and masking. *Family Medicine. 39 (2)*, 132-137

Wald, D. R., Gray, B. E. & Eatough, E. M. (2019). Survey and web research. In P. Brough (Ed). *Advanced Research Methods for Applied Psychology, Design Analysis and Reporting*. Routledge: London and New York

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology, 12, 129-140.

Weiss, D. J. & Shanteau, J. (2014). Who's the Best? A Relativistic View of Expertise. *Applied Cognitive Psychology*. *1-11* DOI: 10.1002/acp.3015

Weiss, D. J. & Shanteau, J. (2003). Empirical Assessment of Expertise. *The Journal of the Human Factors and Ergonomics Society. 45(1):* 104-16

Willingham, D. B., Wells, L. A., Farrell, J. M. & Stemwedel, M. E. (2000). Implicit motor sequence learning is represented in response locations. Memory & Cognition, 2 8, 366–375.

Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. Memory & Cognition, 2 7, 561–572.

Willis, G. B. (2005). Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA: Sage Publications

Wolf, L. J., von Hecker, U., Maio, G. R. (2017). Affective and cognitive orientations in intergroup perception. *Personality and Social Psychology Bulletin*. *43*, 828-844.

Worland, J. (2015, 17 April). Why Some Cancer Patients Are Losing Their Fingerprints. *Time*. Retrieved from www.time.com

Wreally (2019). Transcribe. Accessed at https://transcribe.wreally.com/

Wynne, B. (1996). May the Sheep Safely Graze? A reflexive View of the Expert-Lay Knowledge Divide. In S. Lash, B. Szerszynski & B. Wynne (Eds.) "Risk, Environment & Modernity. Towards a New Ecology". SAGE Publications: London

Yoon, S., Feng, J. & Kain, A. K. (2011). Altered Fingerprints: Analysis and Detection. *Proceedings of the 20th ICPR.* Retrieved from http://biometrics.cse.msu.edu/Publications/Fingerprint/YoonFengJain_AlteredFP_AnalysisDetection_PAMI11_a.pdf

Zukier, H. (1982). The dillution effect: The role of correlation and dispersion of predictor variables in the use of nondiagnostic information. *Journal of Personality and Social Psychology*. *43*, 1163-1174

APPENDIXES

Appendix A | Interview Schedule

Three categories of probes were organized. The first one related to the demographics of the subjects in order to make the subject more comfortable by talking for a moment about something that would not ask for their opinions or feelings. A second part related to the topic of the ACE-V and the examiners' work and finally a last moment about their motivation and feelings regarding their work.

Demographics
How many years do you have of experience in fingerprint analysis?
Are you engaged with forensic societies?
How was your training to be a fingerprint expert?
Verification phase of ACE-V process
Can you describe the verification phase of the ACE-V process that is carried in your bureau?
Motivation and work
Do you believe that contextual information influences your motivation?
How does your motivation get influenced by the methodology you need to follow (i.e. following a holistic approach versus a numerical approach)?
What are the influences in your motivation regarding the level of accreditation that your bureau needs to present?
Core motivation
What motivates you to do your job?

Appendix B | Need for Cognition Scale (Cacioppo, Petty & Kao, 1984)

Instructions:

For each of the statements below, please indicate to what extent the statement is characteristic of you. If the statement is extremely uncharacteristic of you (not at all like you) please write a "1" to the left of the question; if the statement is extremely characteristic of you (very much like you) please write a "5" next to the question. Of course, a statement may be neither extremely uncharacteristic nor extremely characteristic of you; if so, please use the number in the middle of the scale that describes the best fit.

Please keep the following scale in mind as you rate each of the statements below:

Extremely uncharacteristic	Somewhat uncharacteristic	ncertain	Somewhat characteristic	Extremely characteristic
1	2	3	4	5

Statement	Answer
1. I would prefer complex to simple problems.	
2. I like to have the responsibility of handling a situation that requires a lot of thinking.	
3. Thinking is not my idea of fun. a	
4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities?	
5. I try to anticipate and avoid situations where there is a likely chance I will have to think in-depth about something."	
6. I find satisfaction in deliberating hard and for long hours.	
7. I only think as hard as I have to.	
8. I prefer to think about small, daily projects to long-term ones?	
9. I like tasks that require little thought once I've learned them?	
10. The idea of relying on thought to make my way to the top appeals to me.	
11. I really enjoy a task that involves coming up with new solutions to problems.	
12. Learning new ways to think doesn't excite me very much?	
13. I prefer my life to be filled with puzzles that I must solve.	
14. The notion of thinking abstractly is appealing to me.	
15. I would prefer a task that is intellectual, difficult, and important to one that	

	is somewhat important but does not require much thought.	
	16. 1 feel relief rather than satisfaction after completing a task that required a lot of mental effort?	
	17. It's enough for me that something gets the job done; I don't care how or why it works?	
	18. I usually end up deliberating about issues even when they do not affect me personally.	

Appendix C | Work Extrinsic and Intrinsic Motivation Scale (Tremblay et al., 2009)

Instructions:

For the statement "Why Do You Do Your Work?", use the scale below, and please indicate to what extent each of the following items corresponds to the reasons why you are presently involved in your work.

Does not correspond at all		Corresponds moderately			Corresponds exactly	
1	2	3	4	5	6	7

Statement	Answer
1. Because this is the type of work I chose to do to attain a certain lifestyle	
2. For the income it provides me	
3. I ask myself this question, I don't seem to be able to manage the important tasks related to this work	
4. Because I derive much pleasure from learning new things	
5. Because it has become a fundamental part of who I am	
6. Because I want to succeed at this job, if not I would be very ashamed of myself.	
7. Because I chose this type of work to attain my career goals.	
8. For the satisfaction I experience from taking on interesting challenges	
9. Because it allows me to earn money.	
10. Because it is part of the way in which I have chosen to live my life.	
11. Because I want to be very good at this work, otherwise I would be very disappointed.	
12. I don't know why, we are provided with unrealistic working conditions.	
13. Because I want to be a "winner" in life.	
14. Because it is the type of work I have chosen to attain certain important objectives.	
15. For the satisfaction I experience when I am successful at doing difficult tasks.	
16. Because this type of work provides me with security.	
17. I don't know, too much is expected of us.	
18. Because this job is a part of my life	

Informed Consent Form for Computer-Based Experiments Research Activity in Human Factors in Fingerprint Examinations

The following informed consent form is for fingerprint examiners who are currently working in any forensic laboratory and who are willing to participate in the research "Human Factors in Fingerprint Examinations", which is going to be carry with computer-based experiments.

This Informed Consent Form will be presented to the participants before starting the experiment.

FRAMEWORK: This research is part of the INTREPID Forensics Programme which is being held at the University of Leicester. The supervision of this research is currently being done by Dr Lisa Smith (Department of Criminology) and Dr Doug Barrett (Department of Psychology). This research is funded by the European Commission Seventh Framework Programme (FP7/2007/2013) under grant agreement no. 607930.

DESCRIPTION: You are invited to participate in a research study that investigates the effects of cognitive contamination in forensic sciences, namely in fingerprint examinations. In this activity, you will be presented with a set of images. You will be asked to decide either if the images are the same or not.

RISKS: There are no known risks involved in the procedure we will ask you to do in this activity, however you can stop at any point of the activity. If you do have any questions about the activity you can contact Francisco Valente Gonçalves at fvg4@leicester.ac.uk. If you want to know more about this research you can also follow the link of its blog http://www.intrepid-forensics.eu/project-6/.

TIME INVOLVEMENT: Your participation will take approximately 30 to 45 minutes.

SUBJECT'S RIGHTS: If you have read this form and have decided to participate in this experiment, please understand your participation is voluntary and you have the right to withdraw your consent or discontinue participation at any time without penalty. Your individual privacy will be maintained in all published and written data resulting from the study once this activity is anonymous.

If you agree with the above-stated conditions and are willing to participate in the experiment, please press "I Accept" below. By accepting the form, you confirm that you meet the following conditions:

- You are currently working as a fingerprint examiner.
- You have read the above consent form, understood it and you agree to it.
- You want to participate in the above-mentioned experiment.

(I Accept Box) **(I Do Not Accept Box)**

Appendix E | Informed Consent Form for Computer-Based Experiments – laypeople

Informed Consent Form for Computer-Based Experiments Research Activity in Human Factors in Fingerprint Examinations

FRAMEWORK: This research is part of the INTREPID Forensics Programme which is being held at the University of Leicester. The supervision of this research is currently being done by Dr Lisa Smith (Department of Criminology) and Dr Doug Barrett (Department of Psychology). This research is funded by the European Commission Seventh Framework Programme (FP7/2007/2013) under grant agreement no. 607930.

DESCRIPTION: You are invited to participate in a research study that investigates the effects of cognitive contamination in forensic sciences, namely in fingerprint examinations. In this activity, you will be presented with a set of images. You will be asked to decide either if the images are the same or not.

RISKS: There are no known risks involved in the procedure we will ask you to do in this activity, however you can stop at any point of the activity. If you do have any questions about the activity you can contact Francisco Valente Gonçalves at fvg4@leicester.ac.uk. If you want to know more about this research you can also follow the link of its blog http://www.intrepid-forensics.eu/project-6/.

TIME INVOLVEMENT: Your participation will take approximately 20 minutes.

SUBJECT'S RIGHTS: If you have read this form and have decided to participate in this experiment, please understand your participation is voluntary and you have the right to withdraw your consent or discontinue participation at any time without penalty. Your individual privacy will be maintained in all published and written data resulting from the study once this activity is anonymous.

If you agree with the above-stated conditions and are willing to participate in the experiment, please press "I Accept" below.

(I Accept Box) **(I Do Not Accept Box)**

Appendix F | Informed Consent Form for Individual Interviews

Informed Consent Form for Computer-Based Experiments Research Activity in Human Factors in Fingerprint Examinations

FRAMEWORK: This research is part of the INTREPID Forensics Programme which is being held at the University of Leicester. The supervision of this research is currently being done by Dr Lisa Smith (Department of Criminology) and Dr Doug Barrett (Department of Psychology). This research is funded by the European Commission Seventh Framework Programme (FP7/2007/2013) under grant agreement no. 607930.

DESCRIPTION: You are invited to participate in a research study that investigates the effects of cognitive contamination in forensic sciences, namely in fingerprint examinations. In this activity, you will be interviewed by the principal researcher of this research activity.

RISKS: You may disclose some personal information during the interview. However, you are allowed to ask the researcher to delete any statement you disclosed at any time. If you do have any questions about the activity you can contact Francisco Valente Gonçalves at fvg4@leicester.ac.uk. If you want to know more about this research you can also follow the link of its blog http://www.intrepid-forensics.eu/project-6/.

TIME INVOLVEMENT: Your participation will take approximately 20 minutes.

SUBJECT'S RIGHTS: If you have read this form and have decided to participate in this interview, please understand your participation is voluntary and you have the right to withdraw your consent or discontinue participation at any time without penalty. Your individual privacy will be maintained in all published and written data resulting from the study once this activity is anonymous.

If you agree with the above-stated conditions and are willing to participate in the experiment, please sign this document.

Participant print name

_____

Participant signature

_____