The effect of task-irrelevant contextual information on judgement, decision-making, and opinion in forensic odontology identification.

A thesis submitted for the degree of
Doctor of Philosophy at the University of Adelaide

Sher-Lin Chiam

Adelaide Dental School
Faculty of Health and Medical Sciences

October 2022

TABLE OF CONTENTS

List of publications arising from this thesis

[1] Chiam S L, Page M, Higgins D, Taylor J. Validity of forensic odontology identification by comparison of conventional dental radiographs: A scoping review. Science & Justice 2019;59:93–101. https://doi.org/10.1016/j.scijus.2018.08.008.

[2] Chiam S L, Higgins D, Colyvas K, Page M, Taylor J. Interpretation, confidence and application of the standardised terms: Identified, Probable, Possible, Exclude and Insufficient in forensic odontology identification. Science and Justice 2021;61:426–34. https://doi.org/10.1016/j.scijus.2021.02.007.

[3] Chiam S L, Dror IE, Huber CD, Higgins D. The biasing impact of irrelevant contextual information on forensic odontology radiograph matching decisions. Forensic Sci Int 2021;327:110997. https://doi.org/10.1016/j.forsciint.2021.110997.

[4] Chiam S L, Louise J, Higgins D. "Identified", "probable", "possible" or "exclude": The influence of task irrelevant information on forensic odontology identification opinion. Science & Justice 2022;62:461–70. https://doi.org/10.1016/j.scijus.2022.06.002.

Abstract

Non task relevant contextual information can give rise to context bias which has been shown to cognitively influence all comparative forensic science decisions that rely on human judgement. However, whether forensic odontology identification opinions which require human judgement and evaluation are affected by contextual bias is, to date, empirically unverified. This project aims to explore and provide sound empirical evidence regarding the influence of non task relevant contextual information on forensic odontology identification casework outcomes.

The opinion formation process in forensic odontology was found to be an under researched area at the start of this thesis. This deficiency could explain the commonly held assumptions that categories on standardised opinion scales represent decisional confidence in identification and that dental radiographs alone provide sufficient information for definitive identification. A scoping review of the available validation studies using dental radiographs for identification reveals support for sufficient inter individual discriminability in dental radiographs to allow matching and definitive identification decisions. While an analysis of the relationship between category levels and confidence found correlations between the identification categories and decision confidence, it remained uncertain whether this confidence was derived only from probabilistic weight estimates of the evidence or if contextual effects contribute as well. These early foundational findings informed the design of the main experiment used to address the central research question of whether non task relevant case information contextually influences identification conclusions and opinions.

In the main experiment in this project, forensic odontologists and dentists participated in an online web based survey where they formed identification opinions. Participants were required to read contextual case information that either supported or contradicted the true match status of pairs of matching or non matching radiographs which they then compared. Subsequently, they were asked to provide probabilistic estimates of whether the pairs of radiographs were a match or non match, assign a category of identification and state their confidence in their decisions.

The overall findings suggest that strong contextual non relevant case information affected the judgement and evaluation process and concluding category decision.

Additionally, training and experience appear to affect the interpretation of the categories used on the identification scale. The tenet and value of scientific expert opinion require the evaluation to be based only on relevant information. The finding that contextual information biases the opinion provides an added reason and a strong argument for the management of non relevant contextual information. Concomitantly, the finding that the interpretation and assignment decisions are affected by the connotation, granularity, and positions of the terms in the scale implies that different scales cannot be compared directly. Although more research is required, it does suggest that familiarity with the scale is an important factor for its efficient and correct application. Finally, the finding that the expert participants appear to understand the implications of the identification levels in an opinion better than the comparison group despite the different geographical forensic odontology training and practice backgrounds, is pleasing and provides support for the value of training and calibration.

Keywords:
Forensic odontology
Identification
Judgement and decision making
Rating scales
Contextual bias

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Acknowledgement

There are several people I wish to thank for their various contributions to this project.

My thanks and appreciation are given to:

Assoc Prof Denice Higgins for her invaluable guidance, generosity with her knowledge and time, unending patience, and unwavering support and friendship throughout the whole journey of this project. In short, thank you for being the very best of supervisors.

Dr Christian Huber for his supervision, academic input, and encouragement.

Dr Itiel Dror for his expert advice and guidance during the very early stage of this research and his ongoing academic input.

Dr Kim Colyvas for his patience and encouragement and for mentoring me through the complex statistical analysis.

Dr Jennie Louise for her generosity with her time and sharing of expertise. Thank you for the insightful comments and suggestions.

Dr Mark Page for his supervisory input and encouragement in the early stage of this project.

Mr Steven Saunderson for writing the computer codes to allow for many advanced functions in the software used to run the experiment.

Dr Michael Adena for coding and managing the data from the experiment.

Assoc Prof Jane Taylor for her patience, invaluable supervision and mentoring in the early half of this project. Also, for her friendship, unwavering support, encouragement, and guidance through to the end of this project.

I am also sincerely grateful to the participants who gave considerable precious time and effort to participate in the project; without these contributions, this project would certainly not have been possible.

I would like to thank my family, especially mum and Linda, who had to work their schedule around me through this project. Thank you.

Writing these acknowledgements made me realise how fortunate I have been.

Thank you to all who have made this project possible.

Preface

When I started the research for this thesis, I discovered that no direct studies had been conducted on the reasoning and cognitive processes involved in evaluating dental evidence to decide on an identification. While it is not the purpose of this thesis to study factors that influence judgement and decision making only, knowledge of these areas was required to understand and design research to test the effect that irrelevant context information may have on the decision making process or the outcomes. The background knowledge supporting the research in this thesis is drawn from both traditional and applied psychology theories.

In the early part of my candidature, I spent time synthesising this basic psychological knowledge, inferring, and identifying its cross disciplinary parallels and relevance to forensic odontology. I then applied this knowledge to the design of the main experiment of this thesis.

The main body of this thesis consists of two phases of development to answer the central research question of whether non task relevant contextual information biases the expert conclusions and opinions in forensic odontology identification casework.

In Phase I, two fundamental aspects are examined: (1) the use of dental radiographs for evaluation and comparison, and (2) the interpretation by forensic odontologists of the categories on standardised identification scales. The findings from Phase I informed the methodology developed and used to investigate contextual effects reported in Phase II.

The results of these empirical studies are presented as a collection of papers that I have authored and published over the course of my candidature.

Chapter 1

Introduction

Restoring identity to an unknown deceased individual has significant legal and humanitarian implications and is considered a fundamental human right in many countries [1–3]. Comparison of ante and postmortem dental information is a recognised method of human identification and constitutes one of the principal aspects of the scope of practice in forensic odontology [4–6]. This method of human identification is especially useful in a mass disaster situation [7–9] and is recognised by the International Criminal Police Organization (INTERPOL) as a primary scientific method for identification [10].

Forensic odontology is a dental specialty formally defined by the Dental Board of Australia [11] as: "The branch of dentistry that is involved in the examination and evaluation of dental evidence, which may then be presented in the interests of justice." This forensic science speciality applies expert dental knowledge to assist the courts in their deliberations by making available knowledge that is otherwise inaccessible to the layperson, and traditionally has been regarded as being highly trustworthy.

Forensic odontology identification relies heavily on human judgement, like other long trusted disciplines of "traditional comparative" forensic science such as analysis of fingerprints, tool marks, shoe prints and documents [12]. One issue of concern that has been raised regarding these traditional comparative methods is that the implicit evaluation and decision making processes are subjective and based on tacit knowledge built on an empirically unsubstantiated scientific foundation [12]. Identification, in the traditional forensic sciences including forensic odontology, is based on the concept of individualisation and uniqueness [13–15], and the level of believed "uniqueness" in comparative characteristics depends on tacit knowledge accumulated through implicit learning.

Reservations about the veracity of such scientific foundation claims in traditional comparative disciplines were raised as early as 2005 by Saks [16] and later addressed in an authoritative review of the state of forensic science in the United States by the National

Research Council of the National Academy of Sciences (NAS) among other concerns, such as educational policies and organisational affiliations [12]. Subsequently, the President's Council of Advisors on Science and Technology (PCAST) reviewed and reported on the state of practice of feature comparison methods [17]. That review recommended that techniques using feature comparison methods must be evidence based and advocated the "black box" approach to empirically validate expert claims; this is where the emphasis in validation is on the decisional outcome (i.e., accuracy), rather than the intermediate evaluation process because the process may not be explicit or accessible for evaluation [17]. Significantly, both these reports also raised the issue of the impact of human observer bias and other cognitive sources of error. The NAS report found that inadequate attention was paid to cognitive biases and contextual influence which can potentially impact decision making, especially in subjective feature comparison disciplines [12]. While the PCAST report's focus was on validating the scientific foundation of expert claims for feature comparison disciplines, the issue of context bias was also mentioned [17]. This is because context bias arising from exposure to and the unconscious incorporation of non task relevant information may impinge on the reliability of the expert opinion [17,18].

Reliability embodies different concepts in different fields and domains; however, the overarching concept includes accuracy and between and within person consistency [19]. Scientific metrology defines reliability as precision or deviation, which is integral to validity, whereas jurisprudence connotes it as factual accuracy and trustworthiness [19]. Accordingly, when expert opinion is contaminated by contextual information, trustworthiness is called into question and consequently impacts the weight of the evidence more than admissibility. Reliability can be interpreted as measures of "reproducibility" and "repeatability", which are the terms specified by PCAST [17]. These measures, along with accuracy, constitute the concept of the validity of a method [17]. Repeatability or intra rater variability can be attributed to random environmental or internal affective and mood effects, while reproducibility or inter rater variability can arise from differing context effects on an individual's beliefs and thresholds in decision making [20–22]. The value of expertise is to provide knowledge unavailable to lay decision makers by interpreting evidence in context [23–25]. Contextual information that could impact decisions presents in a variety of forms, from personal communication cues, case notes and case reports. The main concern about the introduction of such non task relevant information is the impact on reliability [26].

A clear concept of contextual information is required to test the existence and impact of context bias in research and to effectively manage such contextual information in operational procedures. The definition and restriction of context information to non task relevant information reduces the confusion and the resistance to censorship and management of context information that has been voiced by some [26]. For example, in research into the effect of context information on trauma assessment in anthropology [27], the relevance of background information such as "mass grave and human rights excavations" to the relevance of trauma determination may be open to challenge. Another example is the contextual bias research of forensic pathology opinions [28,29], where the relevance and need for case history and demographic information to the overall assessment of the cases were greatly contested [30–39]. Lack of agreement on the relevance and the definition of contextual information has contributed to the belief that the inclusion of context makes for more accurate information and is necessary for evidence evaluation [39–43]. It is generally believed that activity level forensic decision tasks, for example, evaluating injuries that involve hypothesizing possible causes require more background case information compared to source attribution activity such as forensic odontology identification [44]. Furthermore, for the activity level tasks, it may also be more difficult to obtain unanimous agreement on which information is relevant [44].

One challenge in testing for the impact of context bias in forensic odontology identification is the lack of clarity around what constitutes relevant information. There has never been an actual survey or agreement on relevant information for identification in forensic odontology, unlike in other disciplines such as toxicology [45]. However, it is thought that agreement on what constitutes relevant information for the identification task is less debatable than for the activity level task [44]. Therefore, an inference of what may constitute relevant information could potentially be synthesised from the indirect information found in the body of forensic odontology literature. These foundational aspects and issues are reviewed in detail in Chapter 2, but the important consideration is, if case information is used as contextual information for testing, that only non task relevant information is presented in the case information.

In feature comparison disciplines, contextual information in empirical research testing is often presented as written case information [46]. A suitable research design depends on the discipline and methodology, but it must simulate the typical working environment, as empirical test scenarios foreign to the usual workflow or method will be counterproductive. Covert testing by inserting test cases amongst routine casework is considered the gold standard [47], however, workflows and work environments in some disciplines may restrict the adoption of such a testing approach. Simulated testing with case information presented as vignettes is an excellent and proven method in multiple disciplines for studying behaviours and decisions [48,49]. Vignettes allow for control and manipulation of the various variables, mainly the direction and strength of the suggestions [49,50]. Vignette scenarios evoke multiple types of unconscious cognitive effects which are reviewed in Chapter 2. Important features of the vignettes are the ecological relevance and credibility, and the subtleness to evoke the desired unconscious cognitive effects [49]. When the purpose is overt and explicit, it might result in the Hawthorne effect, which is when behaviour or decision processes are adjusted because of awareness of being tested, which biases the testing [50,51]. In testing cognitive bias, the human factor needs to be considered in the experimental design [47,52]. In fact, in validating any method that relies on human judgement the psychological influence on human performance should be considered [53]. Some examples of such considerations are fatigue, boredom, and the order effect and learning through the practice of test cases [52]. Furthermore, the use of appropriate test participants, for example, using experts rather than students [46], the stimuli used to test the context, and the measures of the decision process and conclusion are also important considerations [46].

Contextual bias may affect the process or the confidence in the decision without an effect on the outcome, or it may also change the resulting opinion [54]. The extent of the influence depends on multiple factors. The context effect has been found to be most prominent when the context suggestions are very strong, the evidence is unclear or ambiguous, and/or the method lacks rule guidance and depends on the experience and knowledge of the expert [55,56]. Methods that use statistical and rule based analysis to quantify evidence evaluation are thought to be more resilient to the outcome effect of contextual bias. However, even such rule based or metric based decision processes can be susceptible to context effects [57,58]. The strength of the context effect may also influence whether both the decision process and outcome are affected or if it is only the decision confidence and process that are affected [56,59]. Furthermore, the impact on the decision outcome can depend not only on context strength, clarity of evidence, and decision method but also on how opinions are presented, for example, if an opinion is expressed using a category rating scale, the number of available categories will affect the choice [60]. This aspect is reviewed in further detail in Chapter 2. Each of these factors have individual cognitive effects that contribute to the overall decision process and the extent of the context effect on the validity of the evidential opinion.

The forensic community responded rapidly and positively to the two significant reviews by NAS and PCAST [12,17], with considerable efforts invested in the validation of techniques across many disciplines [61–73]. The cognitive aspect of forensic science practices or "cognitive forensics" [74] gradually also gained increased attention from the scientific and legal community [75,76], with many articles appearing in both the white and grey literature since the NAS report [28,77–86]. Context effect has been shown to affect both traditional and metric analytical disciplines, including analysis of fingerprints [87], blood splatter analysis [88], pathology [29], anthropology (both morphologic and metric analysis) [58,89,90] mixed profile DNA evidence interpretation [91] and toxicology [92].

The reputation of forensic odontology fared varyingly in the reviews by NAS and PCAST [12,17]. Identification by dental comparison was not mentioned in either report, but bitemark analysis was heavily criticised as being unscientific and unreliable. While a limited number of subsequent studies have looked at the validity of forensic identification methodologies [67,68,73,93], how biases may influence decision making remains an under researched area. In fact, the judgement and decision process in the evaluation of evidence and the cognitive effects of using the categorical scales for the opinion have not received much empirical examination at all. Publication on context bias in forensic odontology is limited to one empirical investigation which examined the contextual emotional effect on bite mark analysis [83], and a single commentary on the potential effect of context information in forensic odontology [94]. To date, no one has empirically investigated whether cognitive bias is a factor when matching antemortem with postmortem dental data for forensic odontology identifications.

Gap analysis

There is a lack of empirical evidence as to whether the forensic odontology identification method is prone to contextual bias, and more specifically the impact of extraneous case information on decision making. Non task relevant case information may induce prior beliefs about the identification, which may cognitively bias the judgement about the likelihood of the identification. Forensic odontology identification should be based on the independent assessment of the probability of concordance in the dental information to ensure validity and reliability [18]. Although the validity of forensic dental identification was not criticised by the NAS [12] or PCAST [17] reports it is nonetheless important to understand the potential impact of cognitive biases to ensure scientific rigour and thus judicial and community confidence in this discipline.

Empirical evidence is also important to help determine whether mitigation strategies need to be employed. The most common mitigation strategy is the censoring or management of case information, which would demand investment into the restructuring of workforce organisation and workflows.

This thesis addresses this current gap in our knowledge by exploring the impact of contextual information, provided as part of case histories, on the judgement and decision confidence and outcomes in forensic odontology identification.

Research questions

Central research question

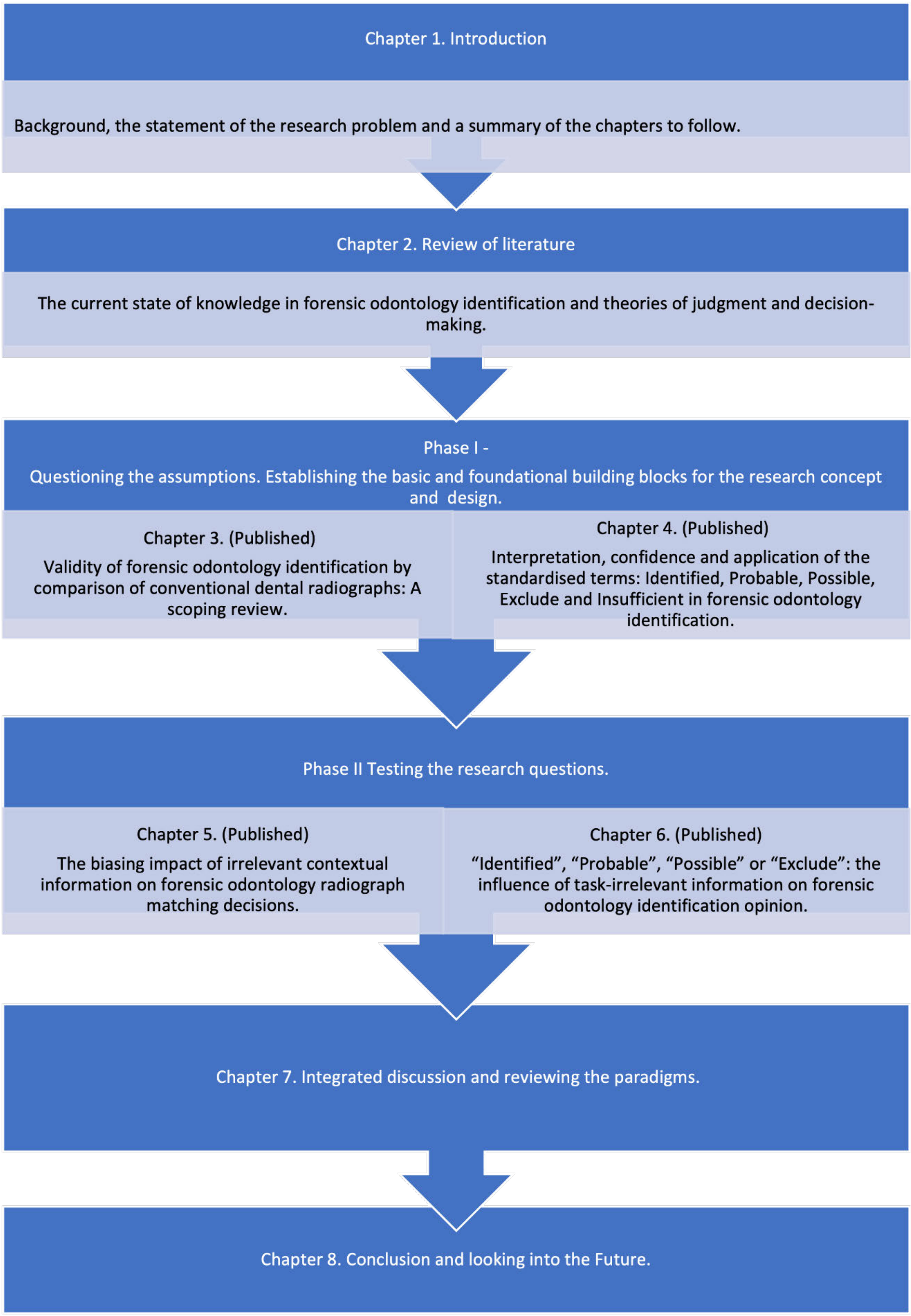
Does the presence of contextual non task relevant information affect the accuracy, concluding opinion and confidence of forensic odontology identification when matching radiographs?

Secondary research questions

1. What does existing research tell us about the validity of using conventional dental radiographs for forensic identification? Can dental radiograph comparison alone allow matching and identification?
2. In the absence of context information, what is the relationship between the selected level of identification (using the Interpol identification scale) and the confidence level when matching dental radiographs of varying levels of difficulty?
3. Does the contextual suggestion of the identity ("Identification" or "Non identification") affect the forensic odontology identification outcome and confidence?
4. Does the strength of the suggestion of identity affect the forensic odontology identification outcome and confidence?
5. Are forensic odontology identification outcomes and confidence affected when the different types of radiographs (from the same person versus different persons) are combined with the different types of context information (strength of suggestion and identity)?
6. Are specialists or experienced practitioners of forensic odontology less affected by contextual information compared to dentists who have no odontology experience? Is there a difference between dentists and laypersons regarding all dependent variables of interest?

Summary and scope of the thesis

The scope of this thesis is presented as a concept map below:



Summary of the thesis

This thesis is presented primarily as a thesis by publication supported by background information to assist the reader.

Chapter 2 reviews the current state of knowledge about forensic odontology identification to provide the reader with a baseline understanding of the principles. It also presents foundation knowledge about cognitive phenomena and the judgement and decision processes that underpin the research presented in this thesis.

Chapter 3 answers the first of the secondary research questions. It presents the first of two publications that represent Phase I of the research which is aimed at questioning the assumptions about the scientific basis for identification utilising forensic odontology. Paper 1 employed the Arksey and O'Malley (2005) [95] methodology for scoping studies to review and summarise the empirical published work on the use of conventional dental radiographs in forensic dental identification from 1990 to 2017. It presents a foundational overview of the methods and discusses the relevant factors that affect validity. This review established that dental radiographs may allow identification to be made without the need for supporting dental information. This level of discriminability means that isolation testing of the effects of contextual information on identification decisions is possible without any confounding factor from relevant information.

Chapter 4 answers the second question in the list of the secondary research questions. It presents the second published paper which further analysed the results from data collected in one of the published papers identified in Chapter 3. The aim was to investigate the relationship of standardised identification categories with self rated confidence levels and binary accuracy assessments. This provided the foundational knowledge and insight into the interpretation and application of the identification category in an identification opinion. This is particularly relevant because, in the reference experiment used in Phase II, decisions were based on identification assessments of dental radiograph pairs only, with no additional dental information provided. Additionally, the radiographs used for the experimental tests in this thesis were selected from the set of radiographs used in the reference study [93], to which contextual information was appended. The main purpose was to examine the application of identification scales in identification decisions and to investigate the long held belief that the scale represents or is a de facto measure of confidence.

With the ground truth established by papers 1 and 2, Phase II tested the hypothesis of the effect of contextual information on the decision making process. The next two chapters address the remainder of the secondary research questions.

Chapters 5 and 6 present the results of the main experiment of this body of work.

Chapter 5 presents the third publication, which introduced the experiment in detail and examined the effect of context information on the accuracy and probability judgement of match (JOM) of practicing forensic odontologists and general dentist participants who were asked to match pairs of dental radiographs supplemented with simulated case information.

Chapter 6 presents the fourth publication, which extended the findings presented in Chapter 5, and examined the effect of context information on decisional choices based on categories of identification used regularly by forensic odontology practitioners globally: “Identified”, “Probable”, “Possible” and “Exclude” (INTERPOL scale version 2009 [96]).

Chapter 7 presents an integrated discussion of findings across the body of research and highlights the relevance of these findings to current practice in forensic odontology.

Chapter 8 concludes this thesis with a summary of the key findings and recommendations of the findings for future research.

References:

[1] Thompson TJU, Black S, British Association for Human Identification. Forensic human identification: an introduction. Print. Boca Raton: CRC Press, ©2007.; 2007.

[2] Rosenblatt A. International Forensic Investigations and the Human Rights of the Dead. Hum Rights Q 2010;32:921–50. https://doi.org/10.1353/hrq.2010.0015.

[3] Humanity after Life: Respecting and Protecting the Dead. ICRC 2019.

[4] Pretty IA, Sweet D. A look at forensic dentistry — Part 1: The role of teeth in the determination of human identity. Br Dent J 2001;190:359–66. https://doi.org/10.1038/sj.bdj.4800972.

[5] Rothwell BR. Principle of dental identification. Dent Clin North Am 2001;45:253–70.

[6] Taylor JA, Kieser JA. Forensic Odontology: Principles and Practice. Wiley Blackwell; 2016. https://doi.org/10.1002/9781118864418.

[7] Taylor J. A brief history of forensic odontology and disaster victim identification practices in Australia. J Forensic Odontostomatol 2009;27:64–74.

[8] Schuller Götzburg P, Suchanek J. Forensic odontologists successfully identify tsunami victims in Phuket, Thailand. Forensic Sci Int 2007;171:204–7. https://doi.org/10.1016/j.forsciint.2006.08.013.

[9] Hill AJ, Hewson I, Lain R. The role of the forensic odontologist in disaster victim identification bushfire. Forensic Sci Int 2011;205:44–7.

[10] INTERPOL Disaster Victim Identification Guide (Version 2014). INTERPOL 2014. https://www.interpol.int/content/download/589/file/18Y1344 E DVI_Guide.pdf (accessed June 19, 2017).

[11] Dental Board of Australia. AUSTRALIAN HEALTH PRACTITIONER REGULATION AGENCY and DENTAL BOARD OF AUSTRALIA 2017. https://www.ahpra.gov.au/Registration/Registers of Practitioners/Specialties and Specialty Fields.aspx (accessed September 28, 2019).

[12] Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council. Strengthening forensic science in the United States: a path forward 2009.

[13] Kieser JA, Bernal V, Neil Waddell J, Raju S, Waddell JN, Raju S, et al. The uniqueness of the human anterior dentition: a geometric morphometric analysis. J Forensic Sci 2007;52:671–7. https://doi.org/10.1111/j.1556 4029.2007.00403.x.

[14] Page M, Taylor J, Blenkin M. Uniqueness in the forensic identification sciences fact or fiction? Forensic Sci Int 2011;206:12–8. https://doi.org/10.1016/j.forsciint.2010.08.004.

[15] Saks MJ. Forensic identification: From a faith based “Science” to a scientific science. Forensic Sci Int 2010;201:14–7. https://doi.org/10.1016/j.forsciint.2010.03.014.

[16] Saks MJ, Koehler JJ. The coming paradigm shift in forensic identification science. Science 2005;309:892–5. https://doi.org/10.1126/science.1111565.

[17] President’s Council of Advisors on Science and Technology, Report to the president Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature Comparison Methods, Washington DC, 2016. https://obamawhite house.archives.gov/sites/default/f 2016.

[18] Thompson WC. What role should investigative facts play in the evaluation of scientific evidence? Australian Journal of Forensic Sciences 2011;43:123–34. https://doi.org/10.1080/00450618.2010.541499.

[19] Stern HS, Cuellar M, Kaye D. Reliability and validity of forensic science evidence. Significance 2019;16:21–4. https://doi.org/10.1111/j.1740 9713.2019.01250.x.

[20] Dror IE, Péron AE, Hind S LL, Charlton D. When emotions get the better of us: The effect of contextual top down processing on matching fingerprints. Appl Cogn Psychol 2005;19:799–809. https://doi.org/10.1002/acp.1130.

[21] Kahneman D, Sibony O, Sunstein CR. Noise: A flaw in human judgment. First edit. New York, NY: Little, Brown Spark; 2021.

[22] Neuberg SL, Judice TN, West SG. What the Need for Closure Scale measures and what it does not: Toward differentiating among related epistemic motives. J Pers Soc Psychol 1997;72:1396–412. https://doi.org/10.1037/0022 3514.72.6.1396.

[23] Evett I. The logical foundations of forensic science: towards reliable knowledge. Philosophical Transactions of the Royal Society B: Biological Sciences 2015;370:20140263. https://doi.org/10.1098/rstb.2014.0263.

[24] Evett IW, Berger CEH, Buckleton JS, Champod C, Jackson G. Finding the way forward for forensic science in the US—A commentary on the PCAST report. Forensic Sci Int 2017;278:16–23. https://doi.org/10.1016/j.forsciint.2017.06.018.

[25] Hicks T, Biedermann A, de Koeijer JA, Taroni F, Champod C, Evett IW. The importance of distinguishing information from evidence/observations when formulating

propositions. Science and Justice 2015;55:520–5. https://doi.org/10.1016/j.scijus.2015.06.008.

[26] National Commission on Forensic Science Human Factors Subcommittee. Ensuring That Forensic Analysis is Based Upon Task Relevant Information. 2015.

[27] Nakhaeizadeh S, Hanson I, Dozzi N. The power of contextual effects in forensic anthropology: a study of biasability in the visual interpretations of trauma analysis on skeletal remains. J Forensic Sci 2014;59:1177–83. https://doi.org/10.1111/1556 4029.12473.

[28] Oliver WR. Effect of History and Context on Forensic Pathologist Interpretation of Photographs of Patterned Injury of the Skin. J Forensic Sci 2017;62:1500–5. https://doi.org/10.1111/1556 4029.13449.

[29] Dror I, Melinek J, Arden JL, Kukucka J, Hawkins S, Carter J, et al. Cognitive bias in forensic pathology decisions. J Forensic Sci 2021;66. https://doi.org/10.1111/1556 4029.14697.

[30] Young TW. Commentary on: Dror IE, Melinek J, Arden JL, Kukucka J, Hawkins S, Carter J, et al. Cognitive bias in forensic pathology decisions. J Forensic Sci. https://doi.org/10.1111/1556 4029.14697. Epub 2021 Feb 20. J Forensic Sci 2021;66:2571. https://doi.org/10.1111/1556 4029.14837.

[31] Speth P, Avedschmidt S, Baeza JJ, Beers DA, Beers KS, Cohle S, et al. Commentary on: Dror IE, Melinek J, Arden JL, Kukucka J, Hawkins S, Carter J, et al. Cognitive bias in forensic pathology decisions. J Forensic Sci. https://doi.org/10.1111/1556 4029.14697. Epub 2021 Feb 20. J Forensic Sci 2021;66:2577–9. https://doi.org/10.1111/1556 4029.14835.

[32] Duflou J. Commentary on: Dror IE, Melinek J, Arden JL, Kukucka J, Hawkins S, Carter J, et al. Cognitive bias in forensic pathology decisions. J Forensic Sci. https://doi.org/10.1111/1556 4029.14697. Epub 2021 Feb 20. J Forensic Sci 2021;66:2561. https://doi.org/10.1111/1556 4029.14836.

[33] Oliver W. Commentary on: Dror IE, Melinek J, Arden JL, Kukucka J, Hawkins S, Carter J, et al. Cognitive bias in forensic pathology decisions. J Forensic Sci. https://doi.org/10.1111/1556 4029.14697. Epub 2021 Feb 20. J Forensic Sci 2021;66:2563–4. https://doi.org/10.1111/1556 4029.14841.

[34] Obenson K. Commentary on: Dror IE, Melinek J, Arden JL, Kukucka J, Hawkins S, Carter J, et al. Cognitive bias in forensic pathology decisions. J Forensic Sci. https://doi.org/10.1111/1556 4029.14697. Epub 2021 Feb 20. J Forensic Sci 2021;66:2582–4. https://doi.org/10.1111/1556 4029.14855.

[35] Peterson BL, Arnall M, Avedschmidt S, Beers D, Bell M, Burton S, et al. Commentary on: Dror IE, Melinek J, Arden JL, Kukucka J, Hawkins S, Carter J, et al. Cognitive bias in forensic pathology decisions. J Forensic Sci. https://doi.org/10.1111/1556 4029.14697. Epub 2021 Feb 20. J Forensic Sci 2021;66:2541–4. https://doi.org/10.1111/1556 4029.14843.

[36] Graber ML. Commentary on: Dror IE, Melinek J, Arden JL, Kukucka J, Hawkins S, Carter J, et al. Cognitive bias in forensic pathology decisions. J Forensic Sci. https://doi.org/10.1111/1556 4029.14697. Epub 2021 Feb 20. J Forensic Sci 2021;66:2574. https://doi.org/10.1111/1556 4029.14857.

[37] Gill JR, Pinneri K, Denton JS, Aiken SS. Commentary on: Dror IE, Melinek J, Arden JL, Kukucka J, Hawkins S, Carter J, et al. Cognitive bias in forensic pathology decisions. J Forensic Sci. https://doi.org/10.1111/1556 4029.14697. Epub 2021 Feb 20. J Forensic Sci 2021;66:2554. https://doi.org/10.1111/1556 4029.14838.

[38] Tse R, Glenn C, Kesha K, Morrow P, Simon S. Commentary on: Dror I, Melinek J, Arden JL, Kukucka J, Hawkins S, Carter J, et al. Cognitive bias in forensic pathology decisions. J Forensic Sci. doi: 10.1111/1556 4029.14697. Epub 2021 Feb 20. J Forensic Sci 2021;66:2567–8. https://doi.org/10.1111/1556 4029.14839.

[39] Oliver WR. Comment on Cognitive bias research in forensic science: A systematic review. Forensic Sci Int 2019;304:35–46. https://doi.org/10.1016/j.forsciint.2019.109898.

[40] Oliver WR. Comment on Kukucka, Kassin, Zapf, and Dror (2017), "Cognitive Bias and Blindness: A Global Survey of Forensic Science Examiners." J Appl Res Mem Cogn 2018;7:161. https://doi.org/10.1016/j.jarmac.2018.01.008.

[41] Thornton JI. Letter to the editor A rejection of "working blind" as a cure for contextual bias. J Forensic Sci 2010;55:1663–1663. https://doi.org/10.1111/j.1556 4029.2010.01497.x.

[42] Thompson WC, Ford S, Gilder JR, Inman K, Jamieson A, Koppl R, et al. Commentary on: Thornton JI. Letter to the editor a rejection of "working blind" as a cure for

contextual bias. J Forensic Sci 2010;55(6):1663. J Forensic Sci 2011;56:562–3. https://doi.org/10.1111/j.1556 4029.2010.01679.x.

[43] Curley LJ, Munro J, Lages M, MacLean R, Murray J. Assessing Cognitive Bias in Forensic Decisions: A Review and Outlook. J Forensic Sci 2020;65:354–60. https://doi.org/10.1111/1556 4029.14220.

[44] Gardner BO, Kelley S, Murrie DC, Dror IE. What do forensic analysts consider relevant to their decision making? Science and Justice 2019;59:516–23. https://doi.org/10.1016/j.scijus.2019.04.005.

[45] Hamnett HJ, Jack RE. The use of contextual information in forensic toxicology: An international survey of toxicologists' experiences. Science and Justice 2019;59:380–9. https://doi.org/10.1016/j.scijus.2019.02.004.

[46] Cooper GS, Meterko V. Cognitive bias research in forensic science: A systematic review. Forensic Sci Int 2019;297:35–46. https://doi.org/10.1016/j.forsciint.2019.01.016.

[47] Towler A, White D, Ballantyne K, Searston RA, Martire KA, Kemp RI. Are forensic scientists experts? J Appl Res Mem Cogn 2018;7:199–208. https://doi.org/10.1016/j.jarmac.2018.03.010.

[48] Spalding NJ, Phillips T. Exploring the use of vignettes: From validity to trustworthiness. Qual Health Res 2007;17:954–62. https://doi.org/10.1177/1049732307306187.

[49] Evans SC, Roberts MC, Keeley JW, Blossom JB, Amaro CM, Garcia AM, et al. Vignette methodologies for studying clinicians' decision making: Validity, utility, and application in ICD 11 field studies. International Journal of Clinical and Health Psychology 2015;15:160–70. https://doi.org/10.1016/j.ijchp.2014.12.001.

[50] Kim J. Scenarios in information seeking and information retrieval research: A methodological application and discussion. Libr Inf Sci Res 2012;34:300–7. https://doi.org/10.1016/j.lisr.2012.04.002.

[51] Dror IE. The Error in "Error Rate": Why Error Rates Are So Needed, Yet So Elusive. J Forensic Sci 2020;65:1034–9. https://doi.org/10.1111/1556 4029.14435.

[52] Martire KA, Kemp RI. Considerations when designing human performance tests in the forensic sciences. Australian Journal of Forensic Sciences 2018;50:166–82. https://doi.org/10.1080/00450618.2016.1229815.

[53] Growns B, Martire KA. Human factors in forensic science: The cognitive mechanisms that underlie forensic feature comparison expertise. Forensic Sci Int 2020;2:148–53. https://doi.org/10.1016/j.fsisyn.2020.05.001.

[54] Dror IE. Letter to editor: On proper research and understanding of the interplay between bias and decision outcomes. Forensic Sci Int 2009;191:17–8. https://doi.org/10.1016/j.forsciint.2009.03.012.

[55] Dror IE, Fraser Mackenzie PAF. Cognitive Biases in Human Perception, Judgment, and Decision Making: Bridging Theory and the Real World. In: Rossmo K, editor. In "Criminal Investigative Failures," Taylor & Francis; 2008, p. 53–67.

[56] Maclean CL, Dror I. A Primer on the Psychology of Cognitive Bias. In: Robertson CT, Kesselheim AS, editors. Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science, and Law. 1st ed., Academic Press; 2016, p. 13–24.

[57] Dror IE, Hampikian G. Subjectivity and bias in forensic DNA mixture interpretation. Science and Justice 2011;51:204–8. https://doi.org/10.1016/j.scijus.2011.08.004.

[58] Hartley S, Winburn AP, Dror IE. Metric forensic anthropology decisions: Reliability and biasability of sectioning point based sex estimates. J Forensic Sci 2022;67. https://doi.org/10.1111/1556 4029.14931.

[59] Dror IE. Cognitive forensics and experimental research about bias in forensic casework. Science & Justice 2012;52:128–30. https://doi.org/10.1016/j.scijus.2012.03.006.

[60] Amoo T, Friedman H. Do Numeric Values Influence Subjects' Responses to Rating Scales? Journal of International Marketing and Marketing Research 2001;26.

[61] Diaz M, Ferrer MA, Impedovo D, Malik MI, Pirlo G, Plamondon R. A perspective analysis of handwritten signature technology. ACM Comput Surv 2019;51. https://doi.org/10.1145/3274658.

[62] Spencer NA, Murray JS. A bayesian hierarchical model for evaluating forensic footwear evidence. Annals of Applied Statistics 2020;14:1449–70. https://doi.org/10.1214/20 AOAS1334.

[63] Langley NR, Meadows Jantz L, McNulty S, Maijanen H, Ousley SD, Jantz RL. Error quantification of osteometric data in forensic anthropology. Forensic Sci Int 2018;287:183–9. https://doi.org/10.1016/j.forsciint.2018.04.004.

[64] Kinoshita Y, Ishihara S. Background population: How does it affect LR based forensic voice comparison? International Journal of Speech, Language and the Law 2014;21:191–224. https://doi.org/10.1558/ijsll.v21i2.191.

[65] Langley NR, Meadows Jantz L, McNulty S, Maijanen H, Ousley SD, Jantz RL. Error quantification of osteometric data in forensic anthropology. Forensic Sci Int. 2018;287:183–9.

[66] Petraco NDK, Shenkin P, Speir J, Diaczuk P, Pizzola PA, Gambino C, et al. Addressing the National Academy of Sciences' Challenge: A Method for Statistical Pattern Comparison of Striated Tool Marks. J Forensic Sci 2012;57:900–11. https://doi.org/10.1111/j.1556 4029.2012.02115.x.

[67] Kirchhoff S, Fischer F, Lindemaier G, Herzog P, Kirchhoff C, Becker C, et al. Is post mortem CT of the dentition adequate for correct forensic identification? Comparison of dental computed tomography and visual dental record. Int J Legal Med 2008;122:471–9. https://doi.org/10.1007/s00414 008 0274 y.

[68] Ruder TD, Thali YA, Rashid SNA, Mund MT, Thali MJ, Hatch GM, et al. Validation of post mortem dental CT for disaster victim identification. Journal of Forensic Radiology and Imaging 2016;5:25–30. https://doi.org/10.1016/j.jofri.2016.01.006.

[69] Mattijssen EJAT, Witteman CLM, Berger CEH, Brand NW, Stoel RD. Validity and reliability of forensic firearm examiners. Forensic Sci Int 2020;307. https://doi.org/10.1016/j.forsciint.2019.110112.

[70] Ulery BT, Hicklin RA, Roberts MA, Buscaglia JA. Data on the inter examiner variation of minutia markup on latent fingerprints. Data Brief 2016;8:158–90. https://doi.org/10.1016/j.dib.2016.04.068.

[71] Ulery BT, Hicklin RA, Roberts MA, Buscaglia JA. Changes in latent fingerprint examiners' markup between analysis and comparison. Forensic Sci Int 2015;247:54–61. https://doi.org/10.1016/j.forsciint.2014.11.021.

[72] Ulery BT, Hicklin RA, Buscaglia J, Roberts MA. Accuracy and reliability of forensic latent fingerprint decisions. Proceedings of the National Academy of Sciences 2011;108:7733–8. https://doi.org/10.1073/pnas.1018707108.

[73] Berketa JW, Hirsch RS, Higgins D, James H. Radiographic recognition of dental implants as an aid to identifying the deceased. J Forensic Sci 2010;55:66–70. https://doi.org/10.1111/j.1556 4029.2009.01226.x.

[74] Dror I. The ambition to be scientific: human expert performance and objectivity. Science & Justice 2013;53:81–2. https://doi.org/10.1016/j.scijus.2013.03.002.

[75] Nakhaeizadeh S, Morgan RM, Dror I. The Emergence of Cognitive Bias in Forensic Science and Criminal Investigations. Br J Am Leg Studies 2015;4:527–54.

[76] Found B. Deciphering the human condition: the rise of cognitive forensics. Australian Journal of Forensic Sciences 2015;47:386–401. https://doi.org/10.1080/00450618.2014.965204.

[77] Zapf PA, Dror IE. Understanding and Mitigating Bias in Forensic Evaluation: Lessons from Forensic Science. Int J Forensic Ment Health 2017;16:227–38. https://doi.org/10.1080/14999013.2017.1317302.

[78] Dror IE, Champod C, Langenburg G, Charlton D, Hunt H, Rosenthal R. Cognitive issues in fingerprint analysis: Inter and intra expert consistency and the effect of a 'target' comparison. Forensic Sci Int 2011;208:10–7. https://doi.org/10.1016/j.forsciint.2010.10.013.

[79] Dror IE, Charlton D, Péron AE. Contextual information renders experts vulnerable to making erroneous identifications. Forensic Sci Int 2006;156:74–8. https://doi.org/10.1016/j.forsciint.2005.10.017.

[80] Canter D, Hammond L, Youngs D. Cognitive bias in line up identifications: The impact of administrator knowledge. Science and Justice 2013;53:83–8. https://doi.org/10.1016/j.scijus.2012.12.001.

[81] van den Eeden CAJ, de Poot CJ, van Koppen PJ. Forensic expectations: Investigating a crime scene with prior information. Science and Justice 2016;56:475–81. https://doi.org/10.1016/j.scijus.2016.08.003.

[82] Osborne NKP, Taylor MC, Zajac R. Exploring the role of contextual information in bloodstain pattern analysis: A qualitative approach. Forensic Sci Int 2016;260:1–8. https://doi.org/10.1016/j.forsciint.2015.12.039.

[83] Osborne NKP, Woods S, Kieser J, Zajac R. Does contextual information bias bitemark comparisons? Sci Justice 2014;54:267–73. https://doi.org/10.1016/j.scijus.2013.12.005.

[84] Kerstholt JH, Paashuis R, Sjerps M. Shoe print examinations: Effects of expectation, complexity and experience. Forensic Sci Int 2007;165:30–4. https://doi.org/10.1016/j.forsciint.2006.02.039.

[85] Kerstholt J, Eikelboom A, Dijkman T, Stoel R, Hermsen R, van Leuven B. Does suggestive information cause a confirmation bias in bullet comparisons? Forensic Sci Int 2010;198:138–42. https://doi.org/10.1016/j.forsciint.2010.02.007.

[86] Nakhaeizadeh S, Dror IE, Morgan RM. Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. Science & Justice 2014;54:208–14. https://doi.org/10.1016/j.scijus.2013.11.003.

[87] Dror IE, Charlton D, Péron AE. Contextual information renders experts vulnerable to making erroneous identifications. Forensic Sci Int 2006;156:74–8. https://doi.org/10.1016/j.forsciint.2005.10.017.

[88] Osborne NKP, Zajac R, Taylor MC. Bloodstain Pattern Analysis and Contextual Bias. In: Jamieson A, Moenssens A, editors. Wiley Encyclopedia of Forensic Science, Chichester, UK: John Wiley & Sons, Ltd; 2015, p. 1–7. https://doi.org/10.1002/9780470061589.fsa1127.

[89] Nakhaeizadeh S, Dror IE, Morgan RM. Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. Science & Justice 2014;54:208–14. https://doi.org/10.1016/j.scijus.2013.11.003.

[90] Nakhaeizadeh S, Hanson I, Dozzi N. The power of contextual effects in forensic anthropology: a study of biasability in the visual interpretations of trauma analysis on skeletal remains. J Forensic Sci 2014;59:1177–83. https://doi.org/10.1111/1556 4029.12473.

[91] Dror IE, Hampikian G. Subjectivity and bias in forensic DNA mixture interpretation. Science & Justice 2011;51:204–8. https://doi.org/10.1016/j.scijus.2011.08.004.

[92] Hamnett HJ, Dror IE. The effect of contextual information on decision making in forensic toxicology. Forensic Sci Int 2020;2:339–48. https://doi.org/10.1016/j.fsisyn.2020.06.003.

[93] Page M, Lain R, Kemp R, Taylor J. Validation studies in forensic odontology – Part 1: Accuracy of radiographic matching. Science & Justice 2018;58:185–90. https://doi.org/10.1016/j.scijus.2017.11.001.

[94] Page M, Taylor J, Blenkin M. Context Effects and Observer Bias Implications for Forensic Odontology. J Forensic Sci 2012;57:108–12. https://doi.org/10.1111/j.1556 4029.2011.01903.x.

[95] Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. International Journal of Social Research Methodology: Theory and Practice 2005;8:19–32. https://doi.org/10.1080/1364557032000119616.

[96] INTERPOL Disaster Victim Identification Guide (Version 2009). INTERPOL 2009:1–55. https://www.cmu.edu/chrs/conferences/eppi/docs/Interpol DVI Guide.pdf (accessed June 15, 2019).

Chapter 2

A review of the literature

This chapter reviews the current state of understanding about identification utilising the forensic odontology method and presents an elemental description of the foundations of judgement and decision making. The cognitive psychology theories of judgement and decision making synthesised from traditional and applied psychological literature are described and illustrated through a range of familiar scenarios and situations relevant to the practice of forensic odontology.

Forensic odontology identification method and process: The Importance of radiographs

Central to identification via dental evidence is the evaluation and comparison of antemortem and postmortem dental data to form an opinion regarding the probability of a common source of the two datasets to support or refute identification [1]. This process requires collating and interpreting antemortem records, often from different treatment providers, and performing a virtual or conventional dental autopsy to document postmortem findings and obtain radiographic images [2,3]. The identification process is a complex multi staged operation; hence the collation, transcription, and interpretation at each stage must be quality controlled so that human factor errors and cognitive contamination do not accumulate and cascade to the final critical stage of the reconciliation of the dental data [4,5].

A dental record is a sample representation of an individual's orofacial profile captured as digital and/or analogue modalities. Dental records can consist of written treatment histories, pre and post treatment photographs, three dimensional dental models, and the various forms of radiographic records[1] [3,6,7]. Radiographic records are believed to allow

---

[1] Conventional analogue and digital dental radiographs can be classified into intraoral and extraoral radiographs, and area of radiographic coverage. Intraoral radiographs that include only the coronal portion of teeth are known as bitewings, while those that include the radicular portion of teeth are known as periapical radiographs. Extraoral radiograph known as an Orthopantomograph allows visualisation of the jaws and teeth and is mainly found only in antemortem radiographic records due to the technicality of obtaining this type of exposure.

more definitive opinions to be made because the amount of observable information exceeds that obtained through a visual clinical examination or photographs [8]. The non abstract nature of radiographs also means that the comparative results are visually demonstrable and verifiable, and hence they convey the similarities or differences more effectively than written descriptions. Furthermore, radiographs are considered more reliable than treatment notes because they are less prone to transcription clerical errors and falsification [8]. Despite these attributes, challenges and difficulties are still present in the comparison task when using dental radiographs. This is primarily because conventional digital and analogue radiographs are 2 dimensional representations of 3 dimensional structures and hence technical errors, or even a slight change in orientation or radiophysical parameters, can change the appearance of structures, making direct comparison of images challenging [9,10]. Considerable clinical experience is necessary to understand when these different presentations originated from the same dentition. Therefore, describing the comparison of radiographs in forensic odontology identification as only a "pattern matching exercise" [8,10] is a vast oversimplification of a highly complex mental process.

Radiograph interpretation, evaluation, and comparison: Cognitive information processing

The ability to recognise, evaluate, and compare perceived greyscale shapes and patterns in radiographs is an extension of an innate ability to recognise everyday objects [11]. Although innate, this skill requires high level cognitive processing since it involves matching pre existing memories with different forms of the same object [11,12]. This perception and processing through pre existing memory or knowledge is known as top down processing [11,13].

A bottom up and top down model of information processing has been described in cognitive psychology, where information processing is conceptualised as the interaction of two sources of information input [13,14]. The bottom up process accepts the raw data as it is without applying any pre existing expectations or experiences to it, whereas the top down process uses existing information to make the new data meaningful (essentially filling in the gaps). Top down processing guides the interpretation and understanding of bottom up information; this form of input is also context driven, such that the general interpretation drives the meaning of the bottom up process [14–16]. An example of this effect is the expectation of observing increased physiological and pathological changes in an individual's postmortem dental findings and radiographs when compared to antemortem radiographs due to the extended passage of time between the creation of the two datasets.

Physiological, pathological, and iatrogenic changes in a dentition can result in differences between antemortem and postmortem radiographs [14,17–19]. Therefore, in cases with long chronological intervals between the creation of antemortem and postmortem data or where significant growth changes and/or development in the dentition have occurred, the comparison process may be more complex. This is particularly so when extensive changes resulting from disease or physiological progression and subsequent complex restorations, or extractions, have occurred resulting in the loss of comparable features [17,19]. In such cases, top down processing with inputs from contextual information, and clinical knowledge and/or experience, may help "fill in the gaps". As a result, it becomes possible to create a coherent narrative and explanation of the perceived concordance and non concordance even when there are no treatment records to substantiate the decision, potentially resulting in inappropriate weight being assigned to the evidence and overconfidence in the expert opinion [14,15]. A critical aspect of such expert judgement is in deciding when the differences detected mean that the possibility of the two data sets representing the same person is significantly reduced or even not possible. Two data sets can be considered to be a "non match" when the number and quality of differences detected exceed the tolerance threshold for the data sets to represent one individual.

Deciding on match and identification: decision thresholds

The threshold concept is used in several evidence accumulation and random walk models to explain the dynamic integration of information during decision making [20–25]. This family of models conceptualises the deliberation process as a competition of "signal and noise" whereby the progress and termination of decision making are associated with sequential or concurrent accumulation of information. Once the threshold for the individual or the situation is reached the deliberation ends and the decision is made [20–25]. A part of this threshold relates to confidence levels [24,26–28], but it is unclear whether confidence accumulates during, or post decision; or whether it is combined or separate from other information [29–31]. It does appear that confidence plays an integral role in the process of evidence accumulation [12,26,32]. The decision threshold is thus likely to be influenced by both external and internal sources of input [11,30,33]. Examples of external input are information from context and dental records, while confidence and the implicit sense of the frequency of occurrence of the compared dental radiographic features are examples of internal input.

The threshold model in fingerprint forensic decision making is supported by the qualitative reflexive study of fingerprint examiners [27]. To date, however, no empirical studies have been conducted to examine the exact judgement and decision processes in forensic odontology. It is therefore unclear whether the threshold model is also applicable to forensic odontology, but existing knowledge allows indirect inference to suggest it may be. Firstly, the concept of using a minimum number of concordant points in forensic odontology identification was modelled on fingerprint comparison; based on the belief of similarities in the comparative process between the two disciplines [34–37]. A significant challenge for both these comparative forensic disciplines is quantifying the implicit nature of perceptual comparison and evaluation [35–37]. This suggests forensic odontology radiograph and fingerprint comparisons use similar judgement processes. Second, specific support for the evidence accumulation process can be inferred from the behavioural response of the participants in a radiograph matching validation study by Wenzel et al. [10]. In this study, participants had the option of committing to a definitive decision or deferring when they were not confident enough to make a decision until all additional radiographs of the same area of comparison were viewed. The accumulation of additional information to aid a match decision bears a resemblance to evidence accumulation models, where the decision is only made when sufficient information and confidence are accumulated.

Discriminability in radiographs: accumulating and weighing the value of the dental features

The types of radiographic features used for dental comparison have been found to affect the level of decisional confidence due to variation in their discriminability value and strength [38–41]. Discriminability is generally believed to be higher in intra coronal compared to extra coronal restorations (i.e., amalgam restorations versus crowns) because of the greater visible inter individual variability [42,43]. Of all the intra coronal interventions, endodontic root fillings [44–46], and radiopaque restorations [42,43] are believed to provide a higher evidential yield. Radiation angulation can greatly affect the shape and appearance of some restorations, limiting their discriminability. Despite this, such restorations still provide compelling comparative radiographic evidence as the shape and form of the restorations are determined by the invasive nature of the disease, and the fact that restorations are bespoke. Consequently, a dentition without restoration limits this confidence since only anatomical variation can be compared [47]. Anatomical features include tooth shape, number, spacing, and orientation, bone height and trabecular pattern [48]. Even without restorations, correct identification is still possible [49], although confidence is often lower. This suggests that radiographs alone can offer sufficient information and discriminability to be useful in the decision making process. The question of whether the amount of information and level of discriminability in a single radiograph is enough to allow definitive identification without the support of dental treatment history information is explored further in Chapter 3 of this thesis.

Establishing this baseline is important because it means that radiographic comparison could be used as an experimental representation of the reconciliation process in forensic odontology identification and allow for the isolation and empirical testing of the influence of contextual information. An expert's ability to integrate and incorporate the gauged probabilistic prevalence of the features of comparative interest is posited in Chapter 3 as the reason for allowing identification by comparison of dental data. Different patterns of comparable dental features occur because disease characteristics occur non independently forming population epidemiological trends [40,50,51]. Since forensic odontology identification has few actual statistical references, this sense of prevalence is implicitly based on professional knowledge and experience [35,41,51]. Statistical references regarding the pattern of tooth presence/absence and restorations for the purpose of identification appear to be limited to one published open access database "OdontoSearch 3.2" [52] which is based on epidemiological data collected from different samples in the United States [53]. This data may not apply to all geographical populations since epidemiological studies attribute disease development and access to intervention to genetic factors, geography, and socioeconomic status**.** Since there is no recognised statistical reference database, the determination of evidentiary weight for individual features relies mainly on implicit statistical learning. This dependence on tacit knowledge combined with the lack of rationale evaluation in reports and specific protocols for forming an opinion means accessing and evaluating the quality and type of information used for decision making is hindered.

Informing forensic odontology Identification: relevant and irrelevant context information

All modes of information can consciously and unconsciously weigh into and serve as a context for decisions, however, not all contextual information about the case is relevant to the task requested [54]. A uniform understanding of what should be considered task and non task relevant is important for managing contextual information as well as for the sound design of experiments to investigate context effects in forensic science [55,56]. Context information in this research refers to information, which is non task relevant, rather than domain relevant, in keeping with the definition of task relevant information used by the U.S. National Commission on Forensic Science (NCFS) [54]. Defining relevance according to task rather than domain broadens the application of the concept while clarifying it further. In a domain, tasks vary in nature, which means the information required will also vary [54,57–61].

Research on what constitutes relevant information for a forensic odontology identification task has not been published, and there is no position statement from any professional group either. While antemortem and postmortem dental records are the obvious relevant information required for identification, the importance of prevalence or population dental trends has never been explicated. However, it is evident that these have been implicitly considered because this has been a tenet for abandoning the use of a minimum number of concordant points to quantify identification levels [38,40–42], and has influenced the development of statistical databases [53]. Consequently, in this research, subject participants were not provided with demographic information, dental treatment histories or radiograph exposure dates in the test cases in the main experiment (presented in Chapters 5 and 6). Censorship of all such information allowed strict isolation and control of confounding factors in the testing of non task relevant contextual information. It is believed that exposure to contextual information unconsciously contaminates and cognitively biases the judgement and decision process [15,62–64]. Cognitive bias in psychology is defined as a systematic deviation from a rational decision, with the rational decision being the "ideal" or "statistically correct decision" in the normative model of decision making [65,66]. From a forensic odontology perspective, cognitive bias connotes a biased treatment of evidence which may result in an incorrect identification [67].

Cognitive bias: theory and relevance in decision making

Cognitive bias may result from the use of heuristics, which are the "rule of thumb"; "mental shortcuts" or "intuitive judgements" that are frequently used in day to day deliberations to make quick and often accurate decisions [65,68]. This ingrained mental ability to use heuristics is believed to be also intuitively applied in experienced and competent expert decisions and is the basis of the "fast and frugal" theory of expertise and heuristics research [69,70]. The ability to make efficient decisions is thought to be the cornerstone of expertise because relevant information can be quickly identified from training or experience [71]. On the other hand, the "heuristics and bias" school of thinking maintains that heuristics account for errors in judgement and decision making [65]. Cognitive bias is thought to account for the error and deviation from the statistically ideal normative model found in real life descriptive decision making models. This interpretation is prevalent in forensic research, where the main measure of the effect of cognitive bias is the deviation from the ideal, which is accuracy [67]. To explain how heuristics contribute to the development of cognitive bias a two system processing model has been used. This model proposes two competing cognitive processing systems: System 1 is intuitive, fast, and launched automatically, whereas system 2 is slower, more analytical, and deliberate

[72,73]. When cognitive processing needs to be more deliberate system 2 may override or modulate system 1. Encompassed within this cognitive process is the integration of incoming information with pre existing internal information, and here the concept of top down and bottom up information processing mentioned earlier in the text is important. Top down information is predominantly processed by system 1 while bottom up information is processed by system 2. System 1, the fast, automatic, and experiential system, relies heavily on heuristics to ease the cognitive load of deciding [12]. The trade off for easing the cognitive load is the probability of cognitive biases [12,65].

Some of the more relevant heuristics and their corresponding biases identified in forensic science literature are outlined below [14,15,65,74,75]:

- **Representative heuristics/bias** is where the probability of an object or event being representative of a class or group is assumed by the similarity of the new object to the characteristics of an already known group, i.e., stereotyping.
- **Availability heuristics/bias** relates more to ease of recall and produces the illusion of importance or frequency of an event by its level of familiarity.
- **Anchoring** heuristics results from an overreliance on an internally set baseline (the anchor) for judgement and comparison.
- **Confirmation bias** is one of the most documented types of cognitive bias in forensic science [76,77]**.** In the psychological literature, it refers to the pursuit or interpretation of evidence in ways that are biased towards pre existing beliefs, expectations, or a hypothesis. It is a complex phenomenon that has been attributed to a combination of cognitive strategies [78,79]. It may be initiated at the information perception stage where confirmation bias will direct the search, influenced by the information's anchoring, representative or availability. When the anchoring and availability heuristics drive the expectancy and persistence of beliefs interesting information may receive undue attention while contradictory information may receive less weight [78]. Confirmation bias is more likely when information is ambiguous or unclear, and when decisions are difficult because context information is used to "fill in the gap" and ease cognitive dissonance [78,79].

In addition to the cognitive biases listed above, several other cognitive phenomena are known to affect the decision process. These include:

- **Priming effects** where subliminal and unconscious "cues" influence the next response and can be induced by images, semantics, or social interactions [12,80,81].
- **Framing effects,** in contrast to priming, affect the entire judgement and decision process depending on whether the same quantity or quality of information is presented in a positive or negative light, and can also be induced by environmental, social, or contextual information [11,76,82].
- **Primacy and recency effects** relate to the order in which information is presented and received [75,78]. The first and last pieces of information encountered are the most impressionable and have an immediate effect on the judgement and decision to follow.

An illustration of cognitive and context effects in routine forensic odontology practice

Context information can induce the various types of cognitive bias and effects presented so far, and many of these can potentially be encountered in routine forensic odontology practice. A simplified exaggerated illustration of cognitive effects at play is presented in a case below.

> *"The police did not suspect foul play for this requested* ***routine*** *identification and* ***confirmation*** *of (Mr XYZ-); an* ***elderly gentleman*** *who was* ***found dead and decomposed in bed*** *by the caseworker* ***at her weekly visit.*** *The* ***family has been contacted,*** *according to the family Mr XYZ had visited ABC Dental surgery****.*** *The available antemortem dental information are handwritten notes and a radiograph from* ***10 years*** *ago. "*
>
> This scenario could represent similar **routine** coronial cases requiring "confirmation of identification". **Availability** heuristics may bias easy recall of media reports of increased social isolation and late discovery of deceased elderly persons, adding to the expectation of "no foul play" and a "routine identification case". Identification requires a presumed individual to whom the postmortem findings are compared, or "**confirmed**". These confirmatory types of cases typically form the vast majority of forensic odontology caseload, and this **base rate anchors the expectation** to identify rather than exclude. **Top-down processing** creates a coherent and cohesive narrative that **frames** the whole approach during the comparison of the single available outdated 10-year-old antemortem record with the postmortem data. In this context, a more confirmatory positive test and search strategy [78,79] may be adopted, which means the evaluation may disproportionately focus on seeking similarities rather than differences between the dental records. Additionally, it is expected that concordances/differences will be given **more/less weight**, so differences will be more readily reconciled as changes due to physiological and disease causes. The result can be "Identified" due to **early or premature cognitive closure** with high confidence based on arguably incomplete dental information [12,24,27,76,90]. Conversely, a contextual suggestion of "non-identification" may induce the opposite confirmation effect.

The research presented in this thesis tests context bias by invoking all the above cognitive phenomena and context effects through experimental vignettes in the main experiment reported in Chapters 5 and 6. The narratives are modelled on actual police circumstantial case reports modified to control the strength and direction of influence by inducing heuristic and cognitive biases. Additionally, supplementary cognitive effects are employed to modulate the primary cognitive effect. For example, the incorporation of relevant radiograph exposure dates to the non test cases, primes expectations of such information for all cases. Furthermore, phrases such as “believed to be XYZ” or “the unknown” in the narrative subtly semantically prime the match or non match suggestions respectively. Primacy and recency effects are used as adjuncts with pertinent information placed at the beginning and end of the vignettes to increase the strength of bias. The vignettes provide the vehicle for evoking the multiple cumulative cognitive effects which are posited to potentially result in changes in the evaluation process and confidence in a decision [33].

Confidence: role, representation, and capture in forensic odontology identification decisions

In judgement and decision making, confidence is an awareness of one’s own cognitive processing or likelihood of accuracy that accompanies the judgement or decision [83]. Importantly, when external sources of verification are absent this internal gauge of accuracy is the only source of feedback and it accumulates as the decision making process progresses [84,85]. It has been proposed that confidence can either integrate cumulatively or be accumulated as a separate collateral entity throughout the process [24,29,31,86]. In either case, confidence serves to guide the termination of and commitment to a decision [24,87]and thus represents the decision threshold [26] that expresses and conveys certainty [88]. In a multi stage judgement and decision task, judgement confidence influences the extent and direction of evidence seeking and carries over to the next stage of decision making [30]. For example, exposure to contextual information before comparing dental records may provide a certain level of confidence about the identity and therefore may determine how dental data is evaluated and compared.

Confidence has a reciprocal relationship with information need and search behaviour [89–91]. Low confidence motivates and guides a search for further information to increase confidence in decision making [24,29,91]. Both the quantity and quality of information are important for the sense of confidence [91]; some examples of factors that determine quality are clarity, coherence of multiple pieces of information [30,92,93], the weight of evidence [93] and the level of cognitive ease [65,94]. This sense of confidence in decisions is a metacognitive and reflexive construct in psychological research which allows self rated confidence measures to serve as proxy indicators of the decision process [94].

Contemporaneous recording and rating of confidence are believed to reflect these metacognitive estimates more accurately than post decisional rating [83]. The main experiment presented in this thesis collected contemporaneous decisional confidence levels because contextual information can alter the decision making process resulting in changes in overall confidence. A rating system is commonly used to describe metacognitive assessments of a decision[83], however, conceptually rated confidence may represent different aspects of uncertainty in different individuals. It may convey either the sense of the correctness of a decision or the certainty of an event [85,95]. The former is a true metacognitive concept rather than a subjective probabilistic estimate and is usually the more common representation [83].

A lack of clarity may also apply to the assumed confidence represented by the categories used in forensic odontology identification scales. Scales like the American Board of Forensic Odontology (ABFO) [96], and INTERPOL for example [97,98] attempt to standardise identification decision outcomes using general guidelines, but they do not incorporate specific protocols to reach these conclusions. In these scales, identification categories are assumed to represent levels of confidence in the strength of the evidence [99]. It remains unclear how evidential strength is computed or what information is used to decide this strength since there is no direct empirical support to determine whether evidential strength consists solely of intuitive probabilistic frequency estimates of the concordances, or if other information is used. A reflexive study, such as that conducted with fingerprint examiners [27] is needed to determine whether metacognitive confidence is also one of these entities. Chapter 4 of this thesis explores this question by analysing the correlation between reported confidence and the identification category assigned to the decision. When investigating context effect, it is important to collect both the self rated metacognitive and probabilistic estimate measures, as decision process changes may be subtle and manifest differently in each of the parameters. By including both, it is theoretically possible to capture those confidence changes induced by context bias that may not be strong enough to cause a net change in the decision outcome or level of identification classification.

Identification scales in forensic odontology; rating scales in decision-making

Due consideration must be given to metrological aspects such as the rating scale utilised to capture decision changes brought on by biases such as context [100,101]. Categorical forensic odontology identification scales share the same fundamental qualities as all rating scales [102,103]. The endpoints of each scale are anchored by definitive categories and the intervening points indicate probabilistic levels of certainty. Examples of these categories include "identification", "probable", "possible", and "exclusion". It has been demonstrated that the granularity or number of such categories and the connotations of the categorical terms influence interpretation and choice [101,104]. Precision and clarity are suggested to increase with the number of categories, the recommended number is between 5 to 11 points [103,104]. However, it is more important that the granularity is appropriate, and that the connotation of each option is meaningful [105]. It is expected that subtle differences, or meaning nuances, are lost when categories are limited. For example, the term "possible identification" in the INTERPOL identification scale is defined as: *"There is nothing that excludes the identity but either PM or AM data or both are minimal*." [98]. From this guideline statement, it would imply the presumed identity is possible but does not detail whether the underlying decision leans more towards identification or exclusion; the term could be interpreted as *"minimal for identification or exclusion" or even "*inconclusive"*.*

There has been no research into identification scales in forensic odontology to establish if practitioners interpret terms differently or if there is a communication gap between practitioners and legal decision makers. This disconnect has been demonstrated, for example with the ABFO bite mark scale (version 2006) [106] when laypersons emphasised "Match" rather than "Reasonable scientific certainty" which ranked higher in certainty and scale position. Adjectival connotation can influence raters' interpretations and hence their choices [107], but it is unclear whether this outweighs the effect of positional ranking on the scale [104]. According to research, both are equally important, and their influence depends on the context in which they are used [104]. It is also possible that the position and connotation of categories can produce unequal psychological distances between points in the scale [104,108]. Very little is known about the interval effect in the forensic odontology scales as there has been no direct research into this. When confidence is an integral part of the decision making process [26,94] and the categories represent confidence, then wider rated confidence ranges should indicate wider psychological distances. Consequently, the psychological interval may be wider if there are two choices before the term “Exclude” for example “Probable” and “Possible”, rather than three choices: “Probable identification”, “Possible Identification”, and “Possible Exclusion”. These different judgement criterion ranges are significant because certain categories may absorb changes in the decision threshold, implying that contextual effect may not necessarily result in a change of category allocation.

“Identified”, “Probable”, “Possible” and “Exclude”: Categorisation and its effect on perception, judgement and decision

When choosing a category in the scale artificial boundaries and abrupt cut off points must be imposed on the data continuum; this also means that decision threshold ranges need to be assigned to each category [109]. As a result, the reconciliation task and overall identification opinion decision process become even more complex [110,111]. In addition to being dependent on the scale design, boundary divisions can also be re calibrated depending on external context conditions, so when decisions are difficult and when the threshold lies between categories, contextual information can influence categorical attribution [112]. As a result, multiple factors influence the boundaries of categories [110,113]. The process of imposing boundaries or categorisation is innate as well as learned and is required for the cognition of information (e.g., perception of colour or speech) [110,112]. The phenomenon of category perception refers to the processing and evaluation of the evidence according to categories [110,114]. Categorical perception has been demonstrated to be applicable in forensic science in the evaluation of fingerprint minutiae [109]. A clearer mental distinction between categories characterises learned categorical perception; in other words, there is a larger and smaller psychological scaling distance between and within the categories, respectively [110]. Using an example from dentistry, the trained eye distinguishes between healthy and diseased mucosa but, in reality, the distinction between the two states is a continuum rather than a binary distinction. Due to limited research in forensic odontology, it is uncertain if the categorisation perception effect is applicable or if additional probabilistic judgement processes occur before a categorical decision is made. If such a mental step does exist, it is also unknown whether this judgement is directly correlated to category selection or if additional information is factored into the category choice. This dearth of knowledge prompted the inclusion of a probability estimate step before choosing an identification category in the experiment outlined in Chapters 5 and 6 of this thesis. Learnt categorical perception may also imply that familiarity with the scale is important, that calibrating across different scales with varying points is not straightforward, and that the same terms on different scales cannot be directly substituted. Chapter 4 incorporates these questions as part of the foundational understanding of the use of identification scales by forensic odontologists.

Defining forensic odontology experts and expertise

As a subspecialty of dentistry, forensic odontology is recognised and registrable in some countries, but can be practiced without any specific postgraduate qualification or registration as part of the scope of dentistry in many parts of the world [115]. The legal qualification for being an expert witness is through "training, study or experience" or possessing "specialised knowledge" [116,117], therefore a general dentist may provide forensic odontology expert opinions as dental training can be deemed to constitute "training or study", and practice to provide "experience" leading to the possession of "specialised knowledge" the recognition of expertise is, therefore, focused on credentials or experience, or the combination of both [116]. Social recognition of expertise differs from scientific validation since the cognitive psychology approach requires empirical validation of claims of superior performance [118]. Thus, the scientific approach distinguishes between possessing expertise and being an expert [116,119]. Specialisation in many medical and dental disciplines is an expansion of the foundational knowledge of that discipline. In the case of forensic odontology, the foundational knowledge required is common to both the forensic odontologist (expert) and the general dentist (non expert). However, as noted at the beginning of this chapter, the value and role of forensic odontology is in the application of such knowledge to aid the courts, and therefore an understanding of the legal implication and the function of such opinions would be expected. This may be the predominant distinguishing feature between forensic odontology and general dental expertise. Available research suggests that training, education, and practical experience are important to expert performance in forensic odontology identification [120,121]. The research presented in Chapter 3 of this thesis supports the premise that forensic odontologists consistently outperform other groups, including general dentists, when making accurate decisions about identification from dental records [10,99,122–124]. However, empirical evidence is still lacking to explain the superior performance of the odontologist group. It is unclear which element of specialist training, knowledge, or comprehension of the legal aspect of the opinion sets forensic odontologists apart from general dentists or more generally, what constitutes a forensic odontology expert. Part of expertise training is the implicit learning and development of schemas and strategies that allow efficient selective focus on important information from a large amount of information from multiple sources [125,126]. Research would suggest that exposure to contextual information will impact decisions in all experts [64,125,127,128], consequently testing for contextual bias in the comparative forensic domain is equivalent to validating expert performance under the influence of contextual information.

Considerations in human performance and cognitive bias research

Validating expertise requires proof of superior performance [67,129,130], and therefore, a suitable comparison group is important [117]. It is also important to ensure that practicing experts are used rather than substitutes, such as trainees [67]. Testing human performance necessitates considering psychology and the human state in addition to experimental design factors [117,129]. Since locating and detecting matches has a different cognitive effect from pairwise comparison, this may include employing the right test technique [67]. Additionally, the order in which information is exposed has an impact on cognition[131]. Reducing the observer effect (Hawthorn effect) is crucial, for cognitive bias research. Since the context effect is subconscious or even unconscious, testing for it in an overt or obvious way can paradoxically result in the observer effect and prejudice the findings [132,133].

The use of vignettes might not remove the awareness of being tested; however, they permit different simulated models of actual situations to be tested, which otherwise would not be possible in the ecological work environment [134,135]. Although these are artificial simulated conditions, high internal consistency in vignettes is possible because of the systematic manipulation of the different independent variables in a controlled and consistent manner [134,136]. As with all human performance research, the human condition needs to be considered, meaning these fabricated vignettes should also be randomised and counterbalanced [129,134]. Equally important, the vignettes must be carefully pretested and refined accordingly. Believability must be balanced with subtle but deliberate manipulation. The vignettes must contain enough information to achieve the test objectives, but they must also be of suitable length to counteract fatigue or boredom [134]. Other human factor considerations include the need to randomise participants and case order; the provision of mental breaks and distraction between test cases to account for the natural variation in individual performance ability, pattern learning, and carry over effect, respectively [129].

In the testing of expert performance, the range of selected cases must reflect the actual cases encountered in practice, including the type and level of difficulty [117,129]. While these considerations are important for validating expertise, for the testing of contextual effects difficult or ambiguous complex cases should be used to maximise testing resources, as the context effect is likely more prominent in such situations [11,137].

This chapter has reviewed the fundamentals of forensic odontology identification methodology and judgement and decision making theories. In reviewing the forensic odontology identification process two important conclusions can be highlighted. First, is the common belief that interpretation and comparison of dental radiographs are the most important and determining components in forming an opinion. Second, is the assumption that forensic odontologists understand and apply the categories in the forensic odontology identification scale uniformly; an assumption that may have stemmed from the notion that there is a straightforward relationship between the identification decisional confidence and the category scale because the lay terminology that expresses certainty is used. The central concern in this thesis is that the implicit deliberation and opinion formation processes are subjective, which can open the identification decision process to the potential effects of contextual bias. Therefore, the common cognitive effects of contextual information are reviewed and the potential of the cognitive phenomena at play in forensic odontology is illustrated. This knowledge is vital because it is applied to reverse engineer this thesis's experimental case information and design.

The following two chapters, Chapters 3 and 4, test the assumptions mentioned above to allow for the subsequent experiment that tests for the effect of contextual information reported in Chapters 5 and 6.

References:

[1] Pereira CP, Santos JC. How to do identify single cases according to the quality assurance from IOFOS. The positive identification of an unidentified body by dental parameters: A case of homicide. J Forensic Leg Med 2013;20:169–73. https://doi.org/10.1016/j.jflm.2012.06.004.

[2] Taylor JA, Kieser JA. Forensic Odontology: Principles and Practice. Wiley Blackwell; 2016. https://doi.org/10.1002/9781118864418.

[3] Sweet D. Forensic dental identification. Forensic Sci Int 2010;201:3–4. https://doi.org/10.1016/j.forsciint.2010.02.030.

[4] Nakhaeizadeh S, Morgan RM, Rando C, Dror IE. Cascading Bias of Initial Exposure to Information at the Crime Scene to the Subsequent Evaluation of Skeletal Remains. J Forensic Sci 2018;63:403–11. https://doi.org/10.1111/1556 4029.13569.

[5] Dror IE, Morgan RM, Rando C, Nakhaeizadeh S. Letter to the Editor — The Bias Snowball and the Bias Cascade Effects: Two Distinct Biases that May Impact Forensic Decision Making. J Forensic Sci 2017;62:832–3. https://doi.org/10.1111/1556 4029.13496.

[6] Pretty IA, Sweet D. A look at forensic dentistry — part 1: The role of teeth in the determination of human identity. Br Dent J 2001;190:359–66.

[7] Brown K a. Procedures for the collection of dental records for person identification. J Forensic Odontostomatol 2007;25:63–4.

[8] Forrest A. Collection and recording of radiological information for forensic purposes. Aust Dent J 2012;57:24–32. https://doi.org/10.1111/j.1834 7819.2011.01658.x.

[9] MacLean DF, Kogon SL, Stitt LW. Validation of Dental Radiographs for Human Identification. J Forensic Sci 1994;39:13705J. https://doi.org/10.1520/JFS13705J.

[10] Wenzel A, Richards A, Heidmann J. Matching simulated antemortem and postmortem radiographs from human skulls by dental students and experts: Testing skills for pattern recognition. J Forensic Odontostomatol 2010;28:5–12.

[11] Dror IE, Fraser Mackenzie PAF. Cognitive Biases in Human Perception, Judgment, and Decision Making: Bridging Theory and the Real World. In: Rossmo K, editor. In "Criminal Investigative Failures," Taylor & Francis; 2008, p. 53–67.

[12] Fraser Mackenzie PAF, Bucht RE, Dror IE. Forensic Judgement and Decision Making. In: Zentall TR, Crowley PH, editors. Comparative Decision Making, Oxford University

Press; 2013, p. 385–415. https://doi.org/10.1093/acprof:oso/9780199856800.003.0038.

[13] Kveraga K, Ghuman A, Bar M. Top down prediction in the cognitive brain. Brain Cogn 2007;65:145–68. https://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted.

[14] Maclean CL, Dror I. A Primer on the Psychology of Cognitive Bias. In: Robertson CT, Kesselheim AS, editors. Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science, and Law. 1st ed., Academic Press; 2016, p. 13–24.

[15] Zapf PA, Dror IE. Understanding and Mitigating Bias in Forensic Evaluation: Lessons from Forensic Science. Int J Forensic Ment Health 2017;16:227–38. https://doi.org/10.1080/14999013.2017.1317302.

[16] Shea N. Distinguishing Top Down from Bottom Up Effects. In: Biggs FS, Matthen M, Stokes D, editors. Perception and Its Modalities. Revised 21, Oxford University Press; 2014, p. 73–92. https://doi.org/10.1093/acprof:oso/9780199832798.003.0004.

[17] Kogan SL, Maclean DF, Kogan SL, Maclean DF. Long Term Validation Study of Bitewing Dental Radiographs for Forensic Identification. J Forensic Sci 1996;41:230–2.

[18] Fridell S, Ahlqvist J. The use of dental radiographs for identification of children with unrestored dentitions. Journal of Forensic Odonto Stomatology 2006;24:42–6.

[19] Gorza L, Mânica S. Accuracy of dental identification of individuals with unrestored permanent teeth by visual comparison with radiographs of mixed dentition. Forensic Sci Int 2018;289:337–43. https://doi.org/10.1016/j.forsciint.2018.06.004.

[20] Donkin C, Brown SD. Response Times and Decision Making. In: Wixted JT, editor. Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2018, p. 1–33. https://doi.org/10.1002/9781119170174.epcn509.

[21] Busemeyer JR, Diederich A. Survey of decision field theory. Math Soc Sci 2002;43:345–70. https://doi.org/10.1016/S0165 4896(02)00016 1.

[22] Ratcliff R, Smith PL, Brown SD, McKoon G. Diffusion Decision Model: Current Issues and History Roger. Trends Cogn Sci 2017;20:260–81. https://doi.org/10.1016/j.tics.2016.01.007.Diffusion.

[23] Busemeyer JR, Townsend JT. Decision field theory: a dynamic cognitive approach to decision making in an uncertain environment. Psychol Rev 1993;100:432–59.

[24] Hausmann D, Lage D. Sequential evidence accumulation in decision making: The individual desired level of confidence can explain the extent of information acquisition. Judgment and Decision Making Journal 2008;3:229–43.

[25] Forstmann BU, Ratcliff R, Wagenmakers E J. Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. Annu Rev Psychol 2016;67:641–66. https://doi.org/10.1146/annurev psych 122414 033645.

[26] Newell BR, Bröder A. Cognitive processes, models and metaphors in decision research. Judgm Decis Mak 2008;3:195–204.

[27] Charlton D, Fraser Mackenzie P a F, Dror IE. Emotional experiences and motivating factors associated with fingerprint analysis. J Forensic Sci 2010;55:385–93. https://doi.org/10.1111/j.1556 4029.2009.01295.x.

[28] Balsdon T, Wyart V, Mamassian P. Confidence controls perceptual evidence accumulation. Nat Commun 2020;11:1753. https://doi.org/10.1038/s41467 020 15561 w.

[29] Navajas J, Bahrami B, Latham PE. Post decisional accounts of biases in confidence. Curr Opin Behav Sci 2016;11:55–60. https://doi.org/10.1016/j.cobeha.2016.05.005.

[30] van den Berg R, Zylberberg A, Kiani R, Shadlen MN, Wolpert DM. Confidence Is the Bridge between Multi stage Decisions. Current Biology 2016;26:3157–68. https://doi.org/10.1016/j.cub.2016.10.021.

[31] Baranski J v., Petrusic WM. Testing architectures of the decision–confidence relation. Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale 2001;55:195–206. https://doi.org/10.1037/h0087366.

[32] Yeung N, Summerfield C. Metacognition in human decision making: Confidence and error monitoring. Philosophical Transactions of the Royal Society B: Biological Sciences 2012;367:1310–21. https://doi.org/10.1098/rstb.2011.0416.

[33] Shadlen MN, Kiani R. Decision Making as a Window on Cognition. Neuron 2013;80:791–806. https://doi.org/10.1016/j.neuron.2013.10.047.

[34] Phillips VM, Scheepers CF. A comparison between fingerprint and dental concordant characteristics. The Journal of Forensic Odonlo Stomatology 1990;8:17–9.

[35] Keiser Nielsen S. Dental identification: certainty vs probability. Forensic Sci 1977;9:87–97.

[36] Stimson PG. Radiology in forensic odontology. Dent Radiogr Photogr 1975;48:51 55,64 65.

[37] Sognnaes RF. Dental science as evidence in court. Int J Forensic Dent 1976;3:14–6.

[38] Acharya AB, Taylor JA. Are a minimum number of concordant matches needed to establish identity in forensic odontology? J Forensic Odontostomatol 2003;21:6–13.

[39] de Villiers C, Phillips VM. Person identification by means of a single unique dental features. J Forensic Odontostomatol 1998;16:17–21.

[40] Gustafson G, Johanson G. The value of certain characteristics in dental identification. Acta OdontScand 1963;20:367–89.

[41] Adams BJ. Establishing personal identification based on specific patterns of missing, filled, and unrestored teeth. J Forensic Sci 2003;48:487–96. https://doi.org/10.1520/jfs2002226.

[42] Zondagh H, Phillips VM. The discrimination potential of radio opaque composite restorations for identification: Part 3. Journal of Forensic Odonto Stomatology 2009;27:27–32.

[43] Phillips VM, Stuhlinger M. The discrimination potential of amalgam restorations for identification: Part 1. Journal of Forensic Odonto Stomatology 2009;27:23–6.

[44] van der Meer DT, Brumit PC, Schrader B a, Dove SB, Senn DR. Root Morphology and Anatomical Patterns in Forensic Dental Identification: A Comparison of Computer Aided Identification with Traditional Forensic Dental Identification*. J Forensic Sci 2010;55:1499–503. https://doi.org/10.1111/j.1556 4029.2010.01492.x.

[45] Forrest AS, Wu HY H. Endodontic imaging as an aid to forensic personal identification. Australian Endodontic Journal 2010;36:87–94. https://doi.org/10.1111/j.1747 4477.2010.00242.x.

[46] Khalid K, Yousif S, Satti A. Discrimination Potential of Root Canal Treated Tooth in Forensic Dentistry. JOURNAL of FORENSIC ODONTO STOMATOLOGY 2016;3434:19–26.

[47] Phillips VM, Stuhlinger M. The discrimination potential of amalgam restorations for identification: Part 2. Journal of Forensic Odonto Stomatology 2009;27:23–6.

[48] Balla SB, Forgie A. Identification by comparison of caries free bitewing radiographs : Impact of observer qualifications and their clinical experience. Forensic Sci Criminol 2017;2:1–5. https://doi.org/10.15761/FSC.1000108.

[49] MacLean DF, Kogon SL, Stitt LW. Validation of dental radiographs for human identification. J Forensic Sci 1994;39:1195–200.

[50] MacFarlane TW, MacDonald DG, Sutherland D a. Statistical problems in dental identification. J Forensic Sci Soc 1974;14:247–52.

[51] Martin de Las Heras S, Valenzuela A, Luna JDD, Bravo M. The utility of dental patterns in forensic dentistry. Forensic Sci Int 2010;195:166.e1 166.e5. https://doi.org/10.1016/j.forsciint.2009.11.004.

[52] Aschheim K, Adams B. OdontoSearch 3.2 2018. http://www.odontosearch.com/index.html (accessed June 26, 2022).

[53] Adams BJ. The diversity of adult dental patterns in the United States and the implications for personal identification. J Forensic Sci 2003;48:497–503.

[54] National Commission on Forensic Science Human Factors Subcommittee. Ensuring That Forensic Analysis is Based Upon Task Relevant Information. 2015.

[55] Gardner BO, Kelley S, Murrie DC, Dror IE. What do forensic analysts consider relevant to their decision making? Science and Justice 2019;59:516–23. https://doi.org/10.1016/j.scijus.2019.04.005.

[56] Hamnett HJ, Jack RE. The use of contextual information in forensic toxicology: An international survey of toxicologists' experiences. Science and Justice 2019;59:380–9. https://doi.org/10.1016/j.scijus.2019.02.004.

[57] Evett I. The logical foundations of forensic science: towards reliable knowledge. Philosophical Transactions of the Royal Society B: Biological Sciences 2015;370:20140263. https://doi.org/10.1098/rstb.2014.0263.

[58] Jackson G, Jones S, Booth G, Champod C, Evett IW. The nature of forensic science opinion—a possible framework to guide thinking and practicce in investigation and in court proceedings. Science & Justice 2006;46:33–44. https://doi.org/10.1016/S1355 0306(06)71565 9.

[59] Evett I w., Jackson G, Lambert JA, McCrossan S. The impact of the principles of evidence interpretation on the structure and content of statements. Science & Justice 2000;40:233–9. https://doi.org/10.1016/S1355 0306(00)71993 9.

[60] Hicks T, Biedermann A, de Koeijer JA, Taroni F, Champod C, Evett IW. The importance of distinguishing information from evidence/observations when formulating

propositions. Science and Justice 2015;55:520–5. https://doi.org/10.1016/j.scijus.2015.06.008.

[61] Cook R, Evett IIW, Jackson G, Jones PJP, Lambert J a. J. A hierarchy of propositions: deciding which level to address in casework. Science & Justice 1998;38:231–9. https://doi.org/10.1016/S1355 0306(98)72117 3.

[62] Dror IE. Cognitive and Human Factors in Expert Decision Making: Six Fallacies and the Eight Sources of Bias. Anal Chem 2020;92:7998–8004. https://doi.org/10.1021/acs.analchem.0c00704.

[63] Dror IE, Kukucka J, Kassin SM, Zapf PA. No One is Immune to Contextual Bias—Not Even Forensic Pathologists. J Appl Res Mem Cogn 2018;7:316–7. https://doi.org/10.1016/j.jarmac.2018.03.005.

[64] Dror IE. Human expert performance in forensic decision making: Seven different sources of bias. Australian Journal of Forensic Sciences 2017;49:541–7. https://doi.org/10.1080/00450618.2017.1281348.

[65] Kahneman D. Thinking fast and slow. First Edit. New York : Farrar, Straus and Giroux,; 2011.

[66] de Martino B, Kumaran D, Seymour B, Dolan RJ. Frames, Biases, and Rational Decision Making in the Human Brain. Science (1979) 2006;313:684–7. https://doi.org/10.1126/science.1128356.

[67] Cooper GS, Meterko V. Cognitive bias research in forensic science: A systematic review. Forensic Sci Int 2019;297:35–46. https://doi.org/10.1016/j.forsciint.2019.01.016.

[68] Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science (1979) 1974;185:1124–31. https://doi.org/10.1126/science.185.4157.1124.

[69] Gigerenzer G. From tools to theories: A heuristic of discovery in cognitive psychology. Psychol Rev 1991;98:254–67. https://doi.org/10.1037/0033 295X.98.2.254.

[70] Gigerenzer G, Goldstein DG. Reasoning the Fast and Frugal Way: Models of Bounded Rationality. Psychol Rev 1996;103:650–69. https://doi.org/10.1037/0033 295X.103.4.650.

[71] Kahneman D, Klein G. Conditions for intuitive expertise: a failure to disagree. Am Psychol 2009;64:515–26. https://doi.org/10.1037/a0016755.

[72] Evans JStBT. In two minds: dual process accounts of reasoning. Trends Cogn Sci 2003;7:454–9. https://doi.org/10.1016/j.tics.2003.08.012.

[73] Stanovich KE, West RF. Individual differences in reasoning: Implications for the rationality debate? Behavioral and Brain Sciences 2000;23:645–65. https://doi.org/10.1017/S0140525X00003435.

[74] Wilke A, Mata R. Cognitive Bias. In: Ramachandran VS, editor. Encyclopedia of Human Behavior, vol. 1. 2nd ed., Elsevier; 2012, p. 531–5. https://doi.org/10.1016/B978 0 12 375000 6.00094 X.

[75] Plous S. The psychology of judgment and decision making. New York, McGraw Hill , Inc; 1993.

[76] Ask K, Granhag PA. Motivational sources of confirmation bias in criminal investigations: the need for cognitive closure. Journal of Investigative Psychology and Offender Profiling 2005;2:43–63. https://doi.org/10.1002/jip.19.

[77] Cognitive Bias Effects Relevant to Forensic Science. The Forensic Science Regulator 2015;October.

[78] Nickerson RS. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. Review of General Psychology 1998;2:175–220. https://doi.org/10.1037/1089 2680.2.2.175.

[79] Klayman J. Varieties of Confirmation Bias. Psychol Learn Motiv, vol. 32, 1995, p. 385–418. https://doi.org/10.1016/S0079 7421(08)60315 1.

[80] Doyen S, Klein O, Pichon C L, Cleeremans A. Behavioral Priming: It’s All in the Mind, but Whose Mind? PLoS One 2012;7:e29081. https://doi.org/10.1371/journal.pone.0029081.

[81] Janiszewski C, Wyer RS. Content and process priming: A review. Journal of Consumer Psychology 2014;24:96–118. https://doi.org/10.1016/j.jcps.2013.05.006.

[82] Neal TMS, Grisso T. The cognitive underpinnings of bias in forensic mental health evaluations. Psychology, Public Policy, and Law 2014;20:200–11. https://doi.org/10.1037/a0035824.

[83] Norman E, Price MC. Measuring consciousness with confidence ratings. In: Overgaard M, editor. Behavioral Methods in Consciousness Research, vol. 15, Oxford University Press; 2015, p. 159–80. https://doi.org/10.1093/acprof:oso/9780199688890.003.0010.

[84] Yeung N, Summerfield C. Metacognition in human decision making: confidence and error monitoring. Philosophical Transactions of the Royal Society B: Biological Sciences 2012;367:1310–21. https://doi.org/10.1098/rstb.2011.0416.

[85] Fleming SM, Frith CD. The Cognitive Neuroscience of Metacognition. vol. 9. Springer; 2014. https://doi.org/10.1146/annurev.clinpsy.1.102803.143959.

[86] Navajas J, Hindocha C, Foda H, Keramati M, Latham PE, Bahrami B. The idiosyncratic nature of confidence. Nat Hum Behav 2017;1:810–8. https://doi.org/10.1038/s41562 017 0215 1.

[87] Ais J, Zylberberg A, Barttfeld P, Sigman M. Individual consistency in the accuracy and distribution of confidence judgments. Cognition 2016;146:377–86. https://doi.org/10.1016/j.cognition.2015.10.006.

[88] Bang D, Fusaroli R, Tylén K, Olsen K, Latham PE, Lau JYF, et al. Does interaction matter? Testing whether a confidence heuristic can replace interaction in collective decision making. Conscious Cogn 2014;26:13–23. https://doi.org/10.1016/j.concog.2014.02.002.

[89] Blais A RR, Thompson MM, Baranski J v. Individual differences in decision processing and confidence judgments in comparative judgment tasks: The role of cognitive styles. Pers Individ Dif 2005;38:1701–13. https://doi.org/10.1016/j.paid.2004.11.004.

[90] Mayseless O, Kruglanski AW. What makes you so sure? Effects of epistemic motivations on judgmental confidence. Organ Behav Hum Decis Process 1987;39:162–83. https://doi.org/10.1016/0749 5978(87)90036 7.

[91] Fleisig D. Adding Information May Increase Overconfidence in Accuracy of Knowledge Retrieval. Psychol Rep 2011;108:379–92. https://doi.org/10.2466/04.11.23.PR0.108.2.379 392.

[92] Koriat A, Lichtenstein S, Fischhoff B, Combs B. Reasons for Confidence. Journal of Experimental Psychology:Human Learning and Memory 1980;6:107–18.

[93] Kvam PD, Pleskac TJ. Strength and weight: The determinants of choice and confidence. Cognition 2016;152:170–80. https://doi.org/10.1016/j.cognition.2016.04.008.

[94] Moran R, Teodorescu AR, Usher M. Post choice information integration as a causal determinant of confidence: Novel data and a computational account. Cogn Psychol 2015;78:99–147. https://doi.org/10.1016/j.cogpsych.2015.01.002.

[95] Navajas J, Hindocha C, Foda H, Keramati M, Latham PE. The idiosyncratic nature of confidence. Nat Hum Behav 2017;1. https://doi.org/10.1038/s41562 017 0215 1.

[96] American Board of Forensic Odontology. ABFO Body Identification Guidelines n.d. http://abfo.org/wp content/uploads/2012/08/ABFO Body ID Information Guidelines Feb 2017.pdf (accessed June 18, 2017).

[97] INTERPOL Disaster Victim Identification Guide (Version 2009). INTERPOL 2009:1–55. https://www.cmu.edu/chrs/conferences/eppi/docs/Interpol DVI Guide.pdf (accessed June 15, 2019).

[98] INTERPOL Disaster Victim Identification Guide (Version 2014). INTERPOL 2014. https://www.interpol.int/content/download/589/file/18Y1344 E DVI_Guide.pdf (accessed June 19, 2017).

[99] Page M, Lain R, Kemp R, Taylor J. Validation studies in forensic odontology – Part 1: Accuracy of radiographic matching. Science & Justice 2018;58:185–90. https://doi.org/10.1016/j.scijus.2017.11.001.

[100] DeCastellarnau A. A classification of response scale characteristics that affect data quality: a literature review. Qual Quant 2018;52:1523–59. https://doi.org/10.1007/s11135 017 0533 4.

[101] Amoo T, Friedman H. Do Numeric Values Influence Subjects' Responses to Rating Scales? Journal of International Marketing and Marketing Research 2001;26.

[102] Tanner JM. The essential characteristics of a rating system. Am J Phys Anthropol 1971;35:339–40. https://doi.org/10.1002/ajpa.1330350307.

[103] Taherdoost H. What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale / Attitude Scale / Likert Scale. International Journal of Academic Research in Management (IJARM) 2019;8:2296–1747.

[104] Friedman HH, Amoo T. Rating the Rating Scales. The Journal of Marketing Management 1999;9:114–23.

[105] Pearse N. Deciding on the scale granularity of response categories of likert type scales: The case of a 21 point scale. Electronic Journal of Business Research Methods 2011;9:159–71. https://doi.org/10.1017/CBO9781107415324.004.

[106] Mcquiston D. Communicating Opinion Evidence in the Forensic Identification Sciences: Accuracy and Impact Communicating Opinion Evidence in the Forensic Identification Sciences : Accuracy and Impact 2016;59.

[107] Martire KA, Watkins I. Perception problems of the verbal scale: A reanalysis and application of a membership function approach. Science & Justice 2015;55:264–73. https://doi.org/10.1016/j.scijus.2015.01.002.

[108] Brown TC, Daniel TC. Scaling of ratings: concepts and methods. USDA Forest Service Research Paper 1990;RM 293:1–24.

[109] Dror IE, Langenburg G. "Cannot Decide": The Fine Line Between Appropriate Inconclusive Determinations Versus Unjustifiably Deciding Not To Decide. J Forensic Sci 2019;64:10–5. https://doi.org/10.1111/1556 4029.13854.

[110] Harnad S. To Cognize is to Categorize. Handbook of Categorization in Cognitive Science, Elsevier; 2017, p. 21–54. https://doi.org/10.1016/B978 0 08 101107 2.00002 6.

[111] Kahneman D, Sibony O, Sunstein CR. Noise: A flaw in human judgment. First edit. New York, NY: Little, Brown Spark; 2021.

[112] Eysenick MW, Keane MT. Objects, Concepts, and Categories. Cognitive Psychology : A Student's Handbook, Taylor & Francis group; 2013, p. 306–34.

[113] Koriat A, Sorka H. The construction of category membership judgments: Towards a distributed model. Elsevier Ltd; 2017. https://doi.org/10.1016/B978 0 08 101107 2.00031 2.

[114] Harnad S. Categorical Perception. Encyclopedia of Cognitive Science, Chichester: John Wiley & Sons, Ltd; 2006. https://doi.org/10.1002/0470018860.s00490.

[115] Dental Board of Australia 2020. https://www.dentalboard.gov.au/Registration Standards/Scope of practice registration standard.aspx (accessed June 21, 2022).

[116] Martire KA, Edmond G. Rethinking expert opinion evidence. vol. 40. 2017.

[117] Towler A, White D, Ballantyne K, Searston RA, Martire KA, Kemp RI. Are forensic scientists experts? J Appl Res Mem Cogn 2018;7:199–208. https://doi.org/10.1016/j.jarmac.2018.03.010.

[118] Gobet F. The Future of Expertise : The Need for a Multidisciplinary Approach. Journal of Expertise 2018;1:1–7.

[119] Thomas RP, Lawrence A. Assessment of Expert Performance Compared Across Professional Domains. J Appl Res Mem Cogn 2018;7:167–76. https://doi.org/10.1016/j.jarmac.2018.03.009.

[120] Soomer H, Lincoln MJ, Ranta H, Penttila A, Leibur E, Penttilä A, et al. Dentists' qualifications affect the accuracy of radiographic identification. J Forensic Sci 2003;48:1121–6. https://doi.org/10.1520/jfs2003142.

[121] Pinchi V, Norelli GA, Caputi F, Fassina G, Pradella F, Vincenti C. Dental identification by comparison of antemortem and postmortem dental radiographs: Influence of operator qualifications and cognitive bias. Forensic Sci Int 2012;222:252–5. https://doi.org/10.1016/j.forsciint.2012.06.015.

[122] Pinchi V, Norelli G A, Caputi F, Fassina G, Pradella F, Vincenti C. Dental identification by comparison of antemortem and postmortem dental radiographs: Influence of operator qualifications and cognitive bias. Forensic Sci Int 2012;222:252–5. https://doi.org/10.1016/j.forsciint.2012.06.015.

[123] Sholl SA, Moody GH. Evaluation of dental radiographic identification: an experimental study. Forensic Sci Int 2001;115:165–9. https://doi.org/10.1016/S0379 0738(00)00305 4.

[124] Pretty IA, Pretty RJ, Rothwell BR, Sweet D. The Reliability of Digitized Radiographs for Dental Identification: A Web Based Study. J Forensic Sci 2003;48:2002032. https://doi.org/10.1520/JFS2002032.

[125] Dror IE, Charlton D. The paradox of human expertise: why experts get it wrong. In: Kapur N, editor. The Paradoxical Brain, Cambridge University Press; 2011. https://doi.org/10.1016/s1355 0306(98)72118 5.

[126] Edmond G, Towler A, Growns B, Ribeiro G, Found B, White D, et al. Thinking forensics: Cognitive science for forensic practitioners. Science & Justice 2017;57:144–54. https://doi.org/10.1016/j.scijus.2016.11.005.

[127] Dror IE, Charlton D. Why experts make errors. J Forensic Identif 2006;56:600–16. https://doi.org/Doi: 10.1016/s1355 0306(98)72118 5.

[128] Dror IE. A hierarchy of expert performance. J Appl Res Mem Cogn 2016;5:121–7. https://doi.org/10.1016/j.jarmac.2016.03.001.

[129] Martire KA, Kemp RI. Considerations when designing human performance tests in the forensic sciences. Australian Journal of Forensic Sciences 2018;50:166–82. https://doi.org/10.1080/00450618.2016.1229815.

[130] Charness N, Tuffiash M. The Role of Expertise Research and Human Factors in Capturing, Explaining, and Producing Superior Performance. Human Factors: The Journal of the Human Factors and Ergonomics Society 2008;50:427–32. https://doi.org/10.1518/001872008X312206.

[131] Dror IE, Kukucka J. Linear Sequential Unmasking–Expanded (LSU E): A general approach for improving decision making as well as minimizing noise and bias. Forensic Sci Int 2021;3:100161. https://doi.org/10.1016/j.fsisyn.2021.100161.

[132] Found B. Deciphering the human condition: The rise of cognitive forensics. Australian Journal of Forensic Sciences 2015;47:386–401. https://doi.org/10.1080/00450618.2014.965204.

[133] Found B. Deciphering the human condition: the rise of cognitive forensics. Australian Journal of Forensic Sciences 2015;47:386–401. https://doi.org/10.1080/00450618.2014.965204.

[134] Evans SC, Roberts MC, Keeley JW, Blossom JB, Amaro CM, Garcia AM, et al. Vignette methodologies for studying clinicians' decision making: Validity, utility, and application in ICD 11 field studies. International Journal of Clinical and Health Psychology 2015;15:160–70. https://doi.org/10.1016/j.ijchp.2014.12.001.

[135] Spalding NJ, Phillips T. Exploring the use of vignettes: From validity to trustworthiness. Qual Health Res 2007;17:954–62. https://doi.org/10.1177/1049732307306187.

[136] Kim J. Scenarios in information seeking and information retrieval research: A methodological application and discussion. Libr Inf Sci Res 2012;34:300–7. https://doi.org/10.1016/j.lisr.2012.04.002.

[137] Edmond G, Tangen JM, Searston RA, Dror IE. Contextual bias and cross contamination in the forensic sciences: the corrosive implications for investigations, plea bargains, trials and appeals. Law, Probability and Risk 2015;14:1–25. https://doi.org/10.1093/lpr/mgu018.

Chapter 3

Publication

Chiam S L, Page M, Higgins D, Taylor J. Validity of forensic odontology identification by comparison of conventional dental radiographs: A scoping review. Science & Justice 2019;59:93–101. https://doi.org/10.1016/j.scijus.2018.08.008.

Statement of Authorship

Title of Paper	Validity of forensic odontology identification by comparison of conventional dental radiographs: A scoping review
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Chiam SL, Page M, Higgins D, Taylor J. Validity of forensic odontology identification by comparison of conventional dental radiographs: A scoping review. Sci Justice. 2019; 59:93–101.

Principal Author

Name of Principal Author (Candidate)	Sher-Lin Chiam		
Contribution to the Paper	Conception and design of the project Acquiring of research data Analysis and interpretation of research Wrote manuscript Acted as corresponding author		
Overall percentage (%)	85%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	30th May 2022

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Mark A. Page		
Contribution to the Paper	Supervised development of work Manuscript evaluation.		
Signature		Date	30th May 2022

Name of Co-Author	Denice Higgins		
Contribution to the Paper	Supervised development of work Helped in data interpretation Manuscript evaluation		
Signature		Date	18th September 2022

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Jane Taylor		
Contribution to the Paper	Supervised development of work Helped in data interpretation Manuscript evaluation and editing		
Signature		Date	30th May 2022

Science & Justice 59 (2019) 93–101

Contents lists available at ScienceDirect

Science & Justice

journal homepage: www.elsevier.com/locate/scijus

Review

Validity of forensic odontology identification by comparison of conventional dental radiographs: A scoping review

Sher-Lin Chiam[a,*], Mark Page[a], Denice Higgins[b], Jane Taylor[a]

[a] School of Health Sciences, University of Newcastle, Australia
[b] Forensic Odontology Unit, The University of Adelaide, Australia

1. Introduction

Identification of the deceased by forensic dental comparison is well accepted as valid and efficient and is one of the primary methods relied upon in disaster victim identification. A vital component of this method of identification is image comparison with dental radiographs, which continues to provide the most valuable source of evidence [1,2]. Radiographs as a graphic record of dental status contain more verifiable information and detail than written descriptions or charts. Human error can lead to inaccuracies in written records but an image provides an irrefutable source of information [2]. Recently, there has been increased focus on quantification of accuracy, reliability and the objectivity of comparative forensic science disciplines [3]. The 2011 President's Council of Advisors on Science and Technology (PCAST) report [3], which focused specifically on the issues pertaining to comparative forensic sciences, identified the need for empirical investigations into validation of methods. While identification via dental comparison, in contrast to bitemark analysis, did not come under direct scrutiny in either the 2009 National Academy of Sciences (NAS) report [4] or the PCAST report [3], it is important that validation studies are conducted on all aspects of forensic odontology work.

To date, only one review of the literature from the period of 1990 to 1994 regarding the validity of radiographs for forensic identification has been undertaken [5]. This review looked specifically at the use of bitewing radiographs and discussed only four studies. Empirical research since then has included additional types of radiographs and variations in method design, making it a heterogeneous group of studies. In this study we undertake a scoping review to provide an overview of existing empirical research with regards to the validity of using dental radiographs for identification. Information from identified relevant studies has been extracted, collated and summarised to present a landscape of the research. The main issues pertaining to research into this method of forensic identification are also discussed.

2. Method

This scoping review employs the method described in Arksen and O'Malley 2005 [6], which requires identification of specific research questions, systematic search and selection of studies, charting of the data and summarisation of the outcomes. No evaluation of the quality of studies is included in scoping studies [6]. The objective of this review accordingly, was to map existing empirical validity research on the use of dental radiographs for identification and to extract, collate, tabulate and summarise the findings to provide a landscape of the research to date.

A search strategy using the Boolean search terms; 'forensic dentistry' or 'forensic odontology' combined with 'dental radiology' and 'identification' was employed to search "Web of Science", "Science Direct" and "Medline" databases, restricted to publications in English and excluding duplicates. Based on the title and abstract relevant articles were identified. After full text reading, a second exclusion was performed based on pre-determined inclusion and exclusion criteria. A second search was performed by hand-searching through the references cited in these articles for suitable studies not found in the first search.

Included were all primary research publications on validity or reliability using conventional (analogue and digital) extra and intra oral radiographs for comparison in forensic dental identification by human observers. The exclusion criteria were studies that used advanced imaging techniques i.e. computerised tomography (CT) or magnetic resonance imaging (MRI). Studies that compared other maxillofacial structures for identification such as sinuses, trabecular bone structure in edentulous subjects or included biometric or automated systems for comparison of dental radiographic structures were also excluded.

The main research question identified was:

What existing research on the validity of using conventional dental radiographs for forensic identification is available?

The supporting and specific research questions were:

1. What type and sample size of dental radiographs were used?
2. How many participants were involved and what was their experience/skill level?
3. What methods were employed for comparison of the radiographs?
4. Was case information provided to aid the comparison of the radiographs?

* Corresponding author.
E-mail address: SherLin.Chiam@uon.edu.au (S.-L. Chiam).

https://doi.org/10.1016/j.scijus.2018.08.008
Received 12 March 2018; Received in revised form 6 August 2018; Accepted 27 August 2018

Table 1
A list of the articles in this review.

Authors	Year	Title	Journal
Borrman et al. [16]	1990	Accuracy in establishing identity by means of intraoral radiographs.	The Journal of Forensic Odonlo-Stomatology
Ekstrom et al. [10]	1993	Accuracy among dentists experienced in forensic odontology in establishing identity.	The Journal of Forensic Odonlo-Stomatology
MacLean et al. [17]	1994	Validation of dental radiographs for human identification. *Journal of forensic sciences.*	Journal of Forensic Sciences
Kogan et al. [18]	1996	Long-term validation study of bitewing dental radiographs for forensic identification.	Journal of Forensic Sciences
Sholl et al. [15]	2001	Evaluation of dental radiographic identication: an experimental study.	Forensic science international
Pretty et al. [12]	2003	The reliability of digitized radiographs for dental identification: a Web-based study.	Journal of forensic sciences
Soomer et al. [11]	2003	Dentists' qualifications affect the accuracy of radiographic identification.	Journal of forensic sciences
Fridell et al. [7]	2006	The use of dental radiographs for identification of children with unrestored dentitions.	The Journal of Forensic Odonlo-Stomatology
Wenzel et al. [14]	2010	Matching simulated antemortem and postmortem dental radiographs from human skulls by dental students and experts: testing skills for pattern recognition.	The Journal of Forensic Odonlo-Stomatology
Pinchi et al. [9]	2012	Dental identification by comparison of antemortem and postmortem dental radiographs: Influence of operator qualifications and cognitive bias.	Forensic science international
Kaur Bhullar et al. [19]	2014	Evaluation of dental expertise with intraoral periapical view radiographs for forensic identification.	Journal of forensic dental sciences
Balla et al.[8]	2017	Identification by comparison of caries free bitewing radiographs: Impact of observer qualifications and their clinical experience	Forensic science and Criminology
Page et al. [13]	2017	Validation studies in forensic odontology – Part 1: Accuracy of radiographic matching	Science and Justice

5. What type of scale was used for decision making?
6. Do these studies all have the same focus?
7. What and how were the results derived and presented?

3. Results

The first search produced a total of 336 articles; 154 from Medline, 90 from Science Direct and 42 from Web of Science. A search of relevant titles and review of the abstracts reduced this to 21 articles. Full text reading narrowed this to 11 articles and two other relevant articles were located through hand searching. A total of 13 publications were identified for this review. These publications are summarised in Table 1. The frequency and types of terms in the title that are semantically and conceptually associated with validity or reliability studies are summarised in Table 2. Three studies [7–9] did not use any specific term in the title that indicated the main research interest was evaluating the use of dental radiographs.

The main theme of all 13 articles was validity or reliability of the method of comparing ante-mortem and post-mortem dental radiographs for identification of deceased persons. The articles were summarised and analysed according to the following parameters:

1. The types and number of radiographs used.
2. Number of participants and their skill level.
3. The method employed for comparison of the radiographs.
4. Presence or absence of case information to aid the comparison of the radiographs.
5. The scale used for decision making.
6. Specific research questions within the general theme of validation pertaining to the use of dental radiographs for forensic identification.
7. Analysis of the results.

Table 2
A summary of terms used in the title that suggest the concept of validity and reliability.

Term used in title	Study
Validity	MacLean et al.; Kogan et al.; Page et al.
Accuracy	Borrman et al.; Ekstrom et al.; Soomer et al.
Evaluate	Sholl et al.; Kaur Bhullar et al.
Reliability	Pretty et al.
Test	Wenzel et al.
No specific terms used	Fridell et al.; Pinchi et al.; Balla et al.

3.1. *The type and number of radiographs used in the studies*

The type and number of dental radiographs used in the different studies is summarised in Table 3. The sample sizes of radiographs ranged from six to 280 pairs. Four studies [10 – 13] used radiographs from actual forensic cases. Page et al. [13] and Pretty et al. [12] used radiographs from cases that had also had the identifications confirmed by DNA. Radiographs taken on dry skulls were utilised by two studies [14,15], while the remaining studies [7–9,16–19] used radiographs from living patients seen in clinical practice.

Three studies [9] scanned analogue radiographs for conversion to digital format. One study [14] used digital intra oral radiographs for comparison to analogue radiographs.

Six studies [7,8,10,16–18] used only bitewing radiographs. One study [19] used only periapical radiographs, while two studies [14,15] used a mixture of bitewing and periapical radiographs. Four studies [9,11–13] used a mixture of extra-oral and intra-oral radiographs.

Two studies [8,16] used only radiographs which were from the same individuals (true positive) for comparison. The other studies utilised radiographs from different individuals (true negative) as well as true positive radiographs for comparison, however only 4 studies [13,14,17,18] used equal numbers of positive and negative radiographs.

3.2. *The method employed for comparison of the radiographs*

Two main comparison strategies were employed in the studies, paired presentation and free matching. Paired presentations are where only one ante- mortem and one post-mortem radiograph are presented and made available for comparison for each case. Free matching allowed participants to identify matched pairs from all the radiographs presented.

Six studies [7–10,15,16] used the free matching method, and seven studies used the paired presentation method as summarised in Table 3. Two studies [8,16] did not include extra non-matching radiographs in their free matching method.

3.3. *Types and numbers of participants in each of these studies*

Information relating to the participants in the studies is summarised in Tables 4 and 5. Participants had varying levels of expertise in radiographic interpretation and comparison ranging from lay persons to forensic odontologists. Forensic odontologists in some studies were referred to as dentists with forensic experience rather than specialist

Table 3
Summary of the radiographs used in all the studies.

Studies	Type of radiographs for comparison	Radiographs			Matching strategy	Sourced from patients	Sourced from forensic cases.	Dry skull	Digital	Analogue	Analogue to digital
		True positive	True negative	Extra - Not paired							
Borrman et al.	BW to BW	60 pairs			Free Matching	✓				✓	
Ekstrom et al.	BW to BW	30 pairs	1 pairs		Free Matching	✓	✓			✓	
MacLean et al.	BW to BW	140 pairs	140 pairs		Paired one on one	✓				✓	
Kogan et al.	BW to BW	100 pairs	100 pairs		Paired one on one	✓				✓	
Sholl et al.	PA to PA & BW to PA	15 pairs		5 extra BW or PA am	Free Matching			✓		✓	
Pretty et al.	Pm PA or BW to Am OPG	7	3		Paired one on one		✓				✓
Soomer et al.	Pm PA or BW to Am OPG	9 cases			Paired one on one		✓				✓
Fridell et al.	BW to BW	30 pairs		Extra 20 BW for am	Free Matching	✓				✓	
Wenzel et al.	BW to BW & Bw to PA	51 pairs	51 pairs		Paired one on one			✓	✓(Am BW)	✓ (Pm BW + PA)	
Pinchi et al.	OPG to multiple PA & BW	16 OPG to 37 Intraoral		5 BW or PA	Free Matching	✓					✓
Kaur Bhullar et al.	PA to PA	6	4		Paired one on one	✓				✓	
Balla et al.	BW to BW	7 pairs			Free Matching	✓				✓	
Page et al.	Pm PA or BW to Am OPG	25 Pairs	25 pairs		Paired one on one		✓				✓

BW- Bitewing radiograph.
PA- Periapical radiograph.
OPG- Orthopanthogram.

Table 4
Summary of the types and number of participants in the studies.

Studies, (Identified by the index in Table 1.)	Forensic	General Dentist	Maxfac radiologists	Layperson	Dental student	Others	Note	Total
Borrman et al.	1		6				Lay person included dental assistants.	7
Ekstrom et al.	17[a]							17
MacLean et al.	1[a]	1			1		One of the dentists had considerable experience	3
Kogan et al.	1		1		1			3
Sholl et al.	9	9			9	9 Dental hygienists		36
Pretty et al.	42[a]	68		44		45 Dentist with less forensic experience	2 experiments. 1st had 155 responses, 2nd had 87 responses.	155 / 87
Soomer et al.	40						Participants from 19 countries.	40
Fridell et al.		5	5					10
Wenzel et al.	2		1		10			13
Pinchi et al.	6	12			20	20		78
Kaur Bhullar et al.		20			20	20		60
Balla et al.		11		5		6	Others were dental postgraduate studies including 2 forensic odontology trainees.	22

(continued on next page)

Table 5
Characteristics of the forensic odontologists used in the studies.

Forensic odontologists with some form of formal recognition, member of society or association.	Sholl et al.; Soomer et al.; Wenzel et al.; Page et al.
Stated by author as forensic odontologists	Borrman et al.; Kogan et al.; Pretty et al.; Wenzel et al.
General dentists with experience in forensic odontology	Ekstrom et al.; MacLean et al.; Page et al.

practitioners.

Table 5 shows the main criteria used to define forensic odontologists in the 10 studies [10–18,20]. In four studies [12,14,16,18], the authors stated that the experimental group was forensic odontologists, no further information was given as to how they came to be thus classified. Three studies [10,13,17] used general dentists with forensic odontology experience, while four studies [9,11,13,15] used forensic odontologists who were formally recognised specialists by relevant boards. It should be noted that the study by Page et al. 2017 [13] included both dentists with odontology experience and recognised specialists as the experts.

3.4. The scale used for assessment

Five types of scale were used as shown in Table 6. The most common decision required of the participants was the binary match or non-match two choice decision scale used in nine studies [8–10,13,15–19]. The other scales were three to five levels of choice. A study by Wenzel et al. [14] used three choices but required the participants to reduce this to match or non- match. Three studies [11 – 13] used multiple decision choices. Page et al. [13] used the binary forced choice in addition to three types of multi- level choice scale. It is interesting to note that Pretty et al. [12] described the five-level choice scale as an ABFO scale, however, the ABFO scale has only four levels of choice [21].

3.5. Presence or absence of case information to aid the comparison of radiographs

As shown in Table 7, four studies [7,11,17,18] provided background information, Soomer et al. [11] stated case treatment information was provided, Maclean et al. [17] presented the dates of the ante-mortem and radiographs while Kogan et al. [18] supplied only dates of the ante-mortem radiographs in the investigation of the effect of time interval

Table 6
Scale used for judgment of match or non-match and for calculating accuracy.

Scale used	Decisions	Studies	Note
Two alternative forced choice	Match or non-match	Borrman et al.; Ekstrom et al.; MacLean et al.; Kogan et al.; Sholl et al.; Pinchi et al.; Kaur Bhullar et al.; Balla et al.; Page et al.	
3 Levels	Certain match, certain non-match, uncertain	Wenzel et al.	3 rounds reduced until certain or uncertain. Binary forced choice used in analysis.
3 Levels	Without doubt, possible probable	Fridell et al.	This is not forced choice as under possible and probable, more than one choice was possible.
4 levels	Positive identification. Possible identification, Insufficient evidence, exclusion.	Soomer et al.[a] Page et al.[b]	ABFO scale used[a] Interpol scale and ABFO scale used[b]
5 levels	(Reasonable medical certainty, probable, possible, exclude, inconclusive)[c] (Positive, Probable, Possible, Exclude and Insufficient Evidence)c	Pretty et al.[d] Page et al.[c]	Pretty et al.[d]- Stated as ABFO scale used. Only ABFO used to plot specificity and sensitivity at different threshold. Page et al.[c]- Used DVISys™ scale.

[a] ABFO- American Board of Forensic Odontology. (www. abfo.org).
[b] Interpol -International Criminal Police Organisation: Positive, Probable, Possible and Exclude.
[c] DVISys™ (DVI System International, Plass Data Software).

Table 7
Studies with information provided and types of information.

Study	Type of information provided
MacLean et al.	Dates for the antemortem and post-mortem radiographs provided.
Kogan et al.	Antemortem radiograph exposure dates, no post-mortem dates provided.
Fridell et al.	Sex and antemortem exposure dates of the antemortem radiographs and age of disappearance provided.
Soomer et al.	Case treatment notes for antemortem and short case history for post-mortem radiographs. No details about what the information consisted of.

between ante-mortem and post-mortem radiograph and Fridell et al. [7] provided date of exposure of ante-mortem radiographs and dates of disappearance in the study of ability to identify children without restorations from dental radiographs.

3.6. Specific research questions and specific area of focus in these studies

All studies had an overarching aim of evaluation of the method of using dental radiographs for identification, however, sub themes and specific areas of interest were found within this body of research. An extraction of these sub-themes is presented in Table 8. Four studies [9,11,17,19] queried the effect of the skill sets, qualifications and experience of the participants, while three studies [10,16,17] explored restorative status of the dentition on the accuracy of identification. Wenzel et al. [14] queried the validity of comparing digital to analogue intra oral radiograph, Fridell et al. [7] investigated if children without restorations could be correctly identified by using dental radiographs. Kogan et al. [18] investigated the effect of temporal interval on identification.

3.7. Validity: accuracy and reliability results

The results of the studies were complex to summarise due to the different approaches to analysis and calculations employed. The use of different types of participants, radiographs and different areas of emphasis of research in this group of studies added further complexity.

Table 9 presents the results of the "expert" group of participants in studies that fulfilled at least one criteria for validation studies. All these studies have confirmed the identities of the radiographs used for matching. The first five studies [13,14,17–19] are studies which used the paired strategy affording measurement of accuracy as sensitivity and specificity. The sensitivity ranged from 0.71 to 0.96 while the specificity ranged from 0.85 to 1. It is important to note that Maclean et al. [17] and Kogan et al. [18] only had three participants.

While the fifth study [12] used paired matching, an area under a ROC (Receiver operating characteristic) curve was presented for accuracy, hence, no mean sensitivity or specificity rate was presented.

The last four studies [7,9,10,15] used a free matching strategy and therefore only accuracy rates were available, the accuracy rates ranged from 88 to 94%. Only two studies Pretty et al. [12] and Ekstrom et al. [10] repeated the trials on the same practitioners. Pretty et al. [12] found that forensic odontologists had the highest repeatability and reproducibility while Ekstrom et al. [10] found that the participants did not make the same mistakes in the repeat trial.

Other significant results included the expert group was found to have higher accuracy with higher specificity in seven studies [7,9,12,14,15,17,19]. In addition, three studies [12,14,15] found practicing experience was correlated to performance. Kogan et al. [5] found that sensitivity decreased after 25 years.

4. Discussion

The PCAST report [3] advocated and provided a framework for establishing foundational validation research in the comparative forensic sciences. It was acknowledged that human decision and judgment was integral to these methods, yet little consideration was given to the human factors in the recommendations for the design of research. The social-behavioural aspect of such research is important as commented by Martire and Kemp 2016 [22]. The following discussion when approached from the perspective of requirements for basic foundational validity and the psychosocial aspect of evaluating human performance, provided some points of interest with regards to the material, method and research design.

The appropriate sample for validity testing according to the PCAST report [3] needs to be representative of those encountered in case work. Earlier studies [5,10,15–17] concentrated on comparison of the most common radiographs available at that time; analogue bitewings. Later studies [9,11–14] investigated the effect of the use of digital dental radiographs because technological advancement has impelled changes in methods. This demonstrates that the forensic odontology community was aware of the need for scientific validation of the method before the queries by the advisory committee. More importantly, it indicates that the forensic odontology community was also aware of the need to re-evaluate techniques as new technologies arise.

A number of the studies examined in this review fall short of certain considerations in the design and method for validation studies. Two studies [10,11] used forensic cases where the identities of radiographs were not independently verified. While the use of forensic cases may add authenticity and relevance, confirmation of identification should be counter checked by another modality i.e. DNA. This is significant because "ground truth" is required for computation of the true positive and negative rates. These two complementary rates are composites in an accuracy rate. As recognised by many authors in this group of studies the inclusion of specificity rates is important because the cost of false positive identification is higher than that of false negative [5,10,14,17,19]. For this reason, inclusion of non-matching rather than the exclusive use of matched radiographs is important as is the use of equal numbers of non-matching and matching radiographs for a balanced study design. Only four studies [5,13,14,17] had this balanced design. Apart from a balanced design, the total number of radiographs in a trial is also an important consideration. Inclusion of too many radiographs lead to failure of participants to complete the trials, likely due to boredom [13] and fatigue [5], which could confound the accuracy rate. This is an example of the reason for the need to take human factors into account. Ultimately it is human decision and judgment that are being validated and calibrated.

Obtaining large sample sizes of participants is an inherent problem

Table 8
Specific area of focus and research questions found in the articles.

Specific research questions	Studies
Comparison of analogue intra oral radiographs with digital intra oral radiographs.	Wenzel et al.
Identification of children without restorations via the use of dental radiographs.	Fridell et al.
The effect of use of digital dental radiographs using a web interface.	Pretty et al.
The effect of long time span between antemortem and post-mortem radiographs.	Kogan et al.
The effect of presence and absence of restorations between antemortem and post-mortem radiographs.	Borrman et al.; Ekstrom et al.; MacLean et al.
Quality of the observers and the effects on the accuracy when using this method.	Soomer et al.; Pinchi et al.; Kaur Bhullar et al.; Balla et al.;

Table 9
Results from the nine studies that partially fulfill the Pcast criteria for a validation study.

Study	Type of radiograph for comparison	Accuracy: Overall accuracy $= \frac{TP + TN}{Match + non\ match}$	Sensitivity- TP rate $= \frac{TP}{Total\ match}$	Specificity- TN rate $= \frac{TN}{Total\ non-match}$	Other methods used for calculating accuracy	Accuracy rate calculated by other method
MacLean et al.	BW to BW	mean = 0.93 as stated	0.71–0.99	0.97–0.99		
Kogan et al.	BW to BW	mean = 0.93	ave = 0.85	ave = 0.99		
Kaur Bhullar et al.	PA to PA	90.5%	89.3% SD ± 13.	92.3% SD ± 12.6		
Page et al.	Mixed OPG, PA and BW	mean accuracy = 87.5%. (mean accuracy = 96.7% when calculated from multi-level decision scale) [a]	TP = 89.3%,	TN85.6%		
Wenzel et al.	BW to BW Bw to PA	No accuracy rate calculated	0.96 TP	TN = 1	Scores presented as table.	
Pretty et al.	Mixed OPG, PA and BW	0.93 – for forensic odontologists			Receiver operator curve (ROC)- Area under Curve	
Fridell et al.	BW to BW				Total correct, no specificity	91% for 6–7 and 94 for 12–13
Ekstrom et al.	BW to BW				Not calculated presented results of 2 trials	0.88 first trial, 0.89 s trial
Sholl et al.	PA to PA BW to PA				Not calculated. Stated in abstract	93.3%- (Stated only in abstract)
Pinchi et al.	OPG to multiple PA and BW	96% = termed correct attribute	0.96	1	Unspecified for accuracy as described in article.	0.97

Study	Statistical method used	Decision scale	Matching design	Case information	Significant notes or conclusion	Repeat trials and notes of those with repeat trials.
MacLean et al.	Overall accuracy, sensitivity and specificity.	Binary forced choice	Paired matching (140 matched and 140 non-matched)	Dates for ante mortem radiographs and post-mortem radiographs	Forensic odontologists higher specificity rate	No
Kogan et al.	Overall accuracy, sensitivity and specificity.	Binary forced choice	Paired matching (100matched and 100 non -matched)	Partial - patient name and antemortem radiograph date	Sensitivity decreased significantly after 25 year interval between radiographs. Most of these older radiographs had significant restorations and changes from tooth lost.	No
Kaur Bhullar et al.	χ test and Mann Whitney *U* test to compare sensitivity and accuracy	Binary forced choice	Paired matching (6 matched and 4 non-matched)	No	Specialist performed better than general dentists. Higher specificity	No
Page et al.	Overall accuracy, sensitivity and specificity.[a]	Binary forced choice	Paired matching (25 matched and 25 unmatched) (verified)	No	Dental knowledge needed. Dental professionals accurate than layman	No
Wenzel et al.	NA	Binary choice after initial 3 choices	Paired matching (51 pairs of match and non-match)	No	Forensic dentists were more accurate and needed less number of trials to reach final decision.	No

Table 9 (*continued*)

Study	Statistical method used	Decision scale	Matching design	Case information	Significant notes or conclusion	Repeat trials and notes of those with repeat trials.
Pretty et al.	Are under ROC curve and Kappa for agreement over 2 trials	5 level ABFO	Paired matching (7 matched and 3 non-matched) (verified)	No	Practical experience of the forensic odontologists correlated with accuracy, and highest in repeatability and reproducibility	Yes. 2 trials. Dentists highly experienced had higher inter and intra- rater agreement; higher specificity scores. Kappa inter 0.89, intra 0.95
Fridell et al.	Fishers exact test and Pearson chi-square test	3 levels – 1. without doubt 2. possible 3. probable	Free matching- (extra 20 for unmatched antemortem)	No	Oral maxillofacial radiologists with dentists for each age group of children	No
Ekstrom et al.		Binary forced choice	Free match- one extra non-matched	No	Greatest number of mistakes in cases without restorations.	Yes. 2 trials. Mistakes not consistent inter and intra-rater. Judgments of difficult is different both times and to that of author.
Sholl et al.	NA	Binary forced choice	Free match- extra non-matched (antemortem)	No information but participants told of the extra "antemortem"	Forensic Odontologists with practical experience did better than those qualifications. Some Oral hygienist did better than dentists.	No
Pinchi et al.	Overall accuracy, sensitivity and specificity.	Binary forced choice	Free matching - with extra non-match. "No match" had to be specified for the non -matching radiographs by participants	No	Experienced forensic odontologists performed better then less experienced or higher qualified forensic odontologists. Better sensitivity rate of students attributed to Hawthorn effect.	No

OPG – panoramic radiograph.
PA- periapical radiograph.
BW- bitewing radiograph.
TP- True positive.
TN- True negative.
[a] Page et al. used both a binary decision and multi- level decision scale for the experiment.

in such research because forensic odontology is a subspecialty in dentistry and therefore the number of qualified or active practitioners is limited and performance of practitioners is the crux of such validation research.

The experimental group considered as the "expert" group of practitioners performed better consistently through all studies, however it is important to define who this "expert" should be as the definition of this group varied between studies. Determination of who is an expert is complex because specialist recognition and registration is both societal and evolving [23]. Dentists who practice forensic odontology identification but are not forensic odontology experts by training or education or recognition should rightly be included in validation studies. Comparison of the performance of "expert" with the "non-expert" is a "white box" study. "White box" studies attempt to provide insights into factors that affect decision making and to unpack the implicit cognitive processes [3]. These studies may provide the foundation for refining the factors that affect the quality of decision and cognitive aspects of judgment and decision in this discipline. Foundational validation studies can incorporate designs to allow "white box" observations. For example, Wenzel et al. [14] employed a strategy of reduction of three decision level choices to binary choices. Inference from the choice behaviour revealed that experts needed less number of trials than novices, reflecting their higher degree of confidence together with higher accuracy in the decisions.

A number of studies within this group included information relevant for decision making when comparing the radiographs [5,7,11,17], while the remaining studies [8–10,12–16,19] adopted the model similar to the design for a primary blackbox study referenced in the PCAST report [24], whereby fingerprint samples were presented for matching without any case history. Similarly, these studies isolated the ability to discriminate match and non-match without additional cognitive input apart from the information from the radiographs. The addition of case history or similar information may confound the results if the purpose is primarily validity as defined by PCAST. The inclusion of additional case information may introduce cognitive bias in the interpretation of the radiographs. It may also confound the enhanced ability of certain groups of subjects to integrate the information when comparing the radiographs with the isolated ability to discriminate matched pairs from non-matched pairs of radiographs. In fact, the effect of contextual information on the decision making process in forensic odontology identification has never been explored.

The contention with such primary validation research is the external validity as expressed by a few authors [10,12,17], mainly in actual cases of identification, other dental information is required and is integrated into the final decision. In addition to the inclusion of case information, some studies [7,11,12,14] used multi-level decision scales and the free matching strategy [7–10,15,16] in an attempt to replicate actual working conditions.

Categorical scales may appear representative of actual practice where opinions are expressed as levels of confidence; however, it requires a resolution to a binary decision for validity studies. If this resolution is not done by the participants, the thresholds between correct and incorrect responses have to be determined by the experimenter for statistical analysis [11–13]. The accuracy rate can also differ from such threshold determination as seen in Page et al. [13] where the accuracy rate was higher when the multi-level scales were used. The use of an experimental threshold also introduces an additional layer of interpretation [25] and may not reflect the actual decisions, hence, information about the process and behaviour is lost. Pretty et al. [12] used the categorical choices as threshold and cut off points summarising and reporting using a ROC curve. While this method removes the bias in the observers and provides an overview of accuracy, ROC curve does not allow obvious access to the sensitivity and specificity rates, this is because the area under a ROC curve provides a summed measurement of the ability to discriminate correct and incorrect decisions [26].

Free matching, instead of paired comparison does not allow calculation of specificity because active exclusions cannot be made unless non-matching radiographs were included and actively excluded as employed by Pinchi et al. [9]. Using free matching without inclusion of non-matching radiographs also potentially increases the number of incorrect decisions, as one wrong decision would mean another wrong decision especially where only true matching radiographs are used. This was recognised by Balla et al. [8] and pointed out in the PCAST report [3].

It is apparent from this review that the research questions, design and methodology of existing studies are diverse. The terms used in the titles provided a sense of the diverse approach to the concept to validating a method. Apart from the diversity in approach, it was noted from Table 8, that a number of studies [5,7,8,10,12,14,16,17,19] focused on specific types of radiographs and conditions for example, the comparison of only bitewings [5,10,16,17], restoration free bitewings [8,17], restoration free radiographs of children [7].

It is interesting to query if particular types of radiographs, population or dental status and conditions warrant specific investigations. Should this be analogous to separate error rates required for partial latent fingerprints differing from high quality fingerprint matching recommended in the PCAST report [3]? This has implications for appropriate inference of the external validity of the research. Laboratory based research with good internal validity is the most viable way of establishing foundational validity, however the external validity and application of these error rates in practice may require some deliberation. In the application of an error rate, it is the posterior probability that is paramount for the end user. To illustrate, the main interest for the decision maker is the probability that the remains are the presumed identity given that the opinion is a match. The error rate from research however is the error rate given a match and non-match; sensitivity and specificity. This is analogous to the probability of having a disease given a positive diagnostic test. This probability depends on the prior probability or the prevalence of the disease, however, in the forensic situation, the "prevalence" is not readily available.

The confidence in the opinion of an identification from comparison of radiographs results from personal beliefs in the probability of similarity in natural and iatrogenic dental characteristics. For example, it would appear that confidence in identification is inherently lower in the absence of restorations and outstanding features. As to whether discrimination rate for restoration free radiographs requires specific validation, as was done by Borrman et al., Ekstrom et al. and Macklean et al. [10,16,17] or if such conditions can be treated as a prior probability for modification of the foundation error rate generated from samples of all types of radiographs may require some deliberation.

Further examples will be dental identification in disaster victim identification. In such situations, a common practice is to compare multiple radiographs to find the right match. This was the reason for free matching designs in studies. There is a possibility that the rate will be significantly different due to cognitive bias and choice behaviour of such matching tasks.

Evett et al. [27], who advocates a logic inference approach to evidence, emphasised that an expert opinion is a personal belief, however, the opinion should be evidence based, and calibrated among peers. Perhaps as a discipline, forensic odontologists would benefit from consensus to the approach, model and framework of validation research.

5. Conclusion

The majority of studies in this scoping review were conducted prior to the NAS and PCAST reports. However, the existence of studies of this nature attests to the previous awareness of the forensic odontology community for the need to establish the scientific foundation of this method of identification. The heterogeneity of methodologies employed with regards to approach and research design in previous studies does not enable the collation of results to allow for meaningful comparison of the conclusions drawn. More homogeneous studies with a degree of agreement by the fraternity as to the model and framework suitable would allow for determination of reliable foundational error rates. It remains uncertain if the types of radiographs used for comparison or circumstances of identification i.e. mass disaster scenario versus single identification warrants specific considerations. If further research establishes that these require specific considerations of error rates, possible benefit might also be gained from some consensus on what specific processes would warrant specific error rates separate from the general error rate. This is significant because continual re-validation of this method will be required as technology advances. For example, the use of computed tomography instead of analogue intra oral radiographs for comparison. This review highlights the need for an agreed framework and model within the discipline to serve as a foundation for studies, which can then be integrated and compared to provide more meaningful results.

Declaration of interest

None.

References

[1] D. Sweet, Forensic dental identification, Forensic Sci. Int. 201 (2010) 3–4.

[2] A.S. Forrest, Collection and recording of radiological information for forensic purposes, Aust. Dent. J. 57 (Suppl. 1) (2012) 24–32.

[3] President's Council of Advisors on Science and Technology, Report to the president Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, https://obamawhite-house.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_foren-sic_science_report_final.pdf, (2016) (Washington DC).

[4] Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council, Strengthening Forensic Science in the United States: A Path Forward, National Academy of Sciences, Washington, DC, 2009.

[5] S.L. Kogon, A review of validation studies of dental bitewing radiographs for forensic identification, J. Can. Soc. Forensic Sci. (1996) 113–117.

[6] H. Arksey, L. O'Malley, Scoping studies: Towards a methodological framework, Int. J. Soc. Res. Methodol. Theory Pract. 8 (2005) 19–32, https://doi.org/10.1080/1364557032000119616.

[7] S. Fridell, J. Ahlqvist, The use of dental radiographs for identification of children with unrestored dentitions, J Forensic Odontostomatol. 24 (2006) 42–46.

[8] S.B. Balla, A. Forgie, Identification by comparison of caries free bitewing radiographs: Impact of observer qualifications and their clinical experience, Forensic Sci. Criminol. 2 (2017) 1–5, https://doi.org/10.15761/FSC.1000108.

[9] V. Pinchi, G.A. Norelli, F. Caputi, Dental identification by comparison of antemortem and postmortem dental radiographs: Influence of operator qualifications and cognitive bias, Forensic Sci. Int. 222 (2012) 252–255.

[10] G. Eksrom, Accuracy among dentists experienced in forensic odontology in establishing identity, J Forensic Odontostomatol. 11 (1993) 45–52.

[11] H. Soomer, M.J. Lincoln, H. Ranta, A. Penttilä, E. Leibur, Dentists' qualifications affect the accuracy of radiographic identification, J. Forensic Sci. 48 (2003) 1121–1126.

[12] I.A. Pretty, R.J. Pretty, B.R. Rothwell, D. Sweet, The reliability of digitized radiographs for dental identification: a Web-based study, J. Forensic Sci. 48 (2003) 1–6.

[13] M. Page, R. Lain, R. Kemp, J. Taylor, Validation studies in forensic odontology – Part 1: Accuracy of radiographic matching, Sci. Justice. (2017), https://doi.org/10.1016/j.scijus.2017.11.001.

[14] A. Wenzel, A. Richards, J. Heidmann, Matching simulated antemortem and postmortem radiographs from human skulls by dental studens and experts: Testing skills for pattern recognition, J Forensic Odontostomatol. 28 (2010) 5–12.

[15] S.A. Sholl, G.H. Moody, Evaluation of dental radiographic identication: an experimental study, Forensic Sci. Int. 115 (2001) 165–169.

[16] H. Borrman, H.-G. Grondahl, Accuracy in establishing identity by means of intraoral radiographs, J. Forensic Odonlo-Stomatology. 8 (1990) 31–35.

[17] D.F. MacLean, S.L. Kogon, L.W. Stitt, Validation of dental radiographs for human identification, J. Forensic Sci. 39 (1994) 1195–1200.

[18] S.L. Kogan, D.F. Maclean, S.L. Kogan, D.F. Maclean, Long-term validation study of bitewing dental radiographs for forensic identification validation study of bitewing dental radiographs for forensic, J. Forensic Sci. 41 (1996) 230–232.

[19] K. Kaur Bhullar, R.S. Bhullar, S. Balagopal, A. Ganesh, M. Raian, Evaluation of dental expertise with intraoral periapical view radiographs for forensic identification, J. Forensic Dent. Sci. 6 (2014) 171–176.

[20] G. Ekstrom, T. Johnsson, H. Borman, Accuracy among dentists experienced in forensic odontology in establishing identity, J. Forensic Odonlo-Stomatology. 11 (1993) 45–52.

[21] ABFO, ABFO Body Identification Guidelines, http://gbforensicservices.com/identification1.html, (2000) , Accessed date: 18 June 2017.

[22] K.A. Martire, R.I. Kemp, Considerations when designing human performance tests in the forensic sciences, Aust. J. Forensic Sci. (2016) 1–17.

[23] I.E. Dror, D. Charlton, The paradox of human expertise: why experts get it wrong, in: N. Kapur (Ed.), The Paradoxical Brain, Cambridge University Press, 2011.

[24] B.T. Ulery, R.A. Hicklin, J. Buscaglia, M.A. Roberts, Accuracy and reliability of forensic latent fingerprint decisions, Proc. Natl. Acad. Sci. 108 (2011) 7733–7738.

[25] S. Halligan, D.G. Altman, S. Mallett, Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: A discussion and proposal for an alternative approach, Eur Radiol. 25 (2015) 932–939.

[26] A.-M. Šimundić, Measures of diagnostic accuracy: Basic definitions, Med. Biol. Sci. 19 (2008) 1–9.

[27] I.W. Evett, C.E.H. Berger, J.S. Buckleton, C. Champod, G. Jackson, Finding the way forward for forensic science in the US—A commentary on the PCAST report, Forensic Sci. Int. 278 (2017) 16–23.

Chapter 4

Publication

Chiam S L, Higgins D, Colyvas K, Page M, Taylor J. Interpretation, confidence and application of the standardised terms: Identified, Probable, Possible, Exclude and Insufficient in forensic odontology identification. Science and Justice 2021;61:426–34. https://doi.org/10.1016/j.scijus.2021.02.007.

Statement of Authorship

Title of Paper	Interpretation, confidence, and application of the standardised terms: Identified, Probable, Possible, Exclude and Insufficient in forensic odontology identification.
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Chiam SL, Higgins D, Colyvas K, Page M, Taylor J. Interpretation, confidence, and application of the standardized terms: Identified, Probable, Possible, Exclude and Insufficient in forensic odontology identification. Elsevier B.V.; 2021; 61:426–34. Available: https://doi.org/10.1016/j.scijus.2021.02.007

Principal Author

Name of Principal Author (Candidate)	Sher-Lin Chiam		
Contribution to the Paper	Conception and design of the project Acquiring of research data Analysis and interpretation of research Wrote manuscript Acted as corresponding author		
Overall percentage (%)	85%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	30th May 2022

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Denice Higgins		
Contribution to the Paper	Supervised development of work Helped in data interpretation Manuscript evaluation and revision		
Signature		Date	18th September 2022

Name of Co-Author	Kim Colyvas		
Contribution to the Paper	Helped in data analysis Helped in interpretation Manuscript evaluation and editing		
Signature		Date	31-May-2022

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Mark Page		
Contribution to the Paper	Manuscript evaluation.		
Signature		Date	30.05.2022

Name of Co-Author	Jane Taylor		
Contribution to the Paper	Supervised development of work Helped in data interpretation Manuscript evaluation and editing		
Signature		Date	30.05.2022

Science & Justice 61 (2021) 426–434

Contents lists available at ScienceDirect

Science & Justice

journal homepage: www.elsevier.com/locate/scijus

Interpretation, confidence and application of the standardised terms: Identified, Probable, Possible, Exclude and Insufficient in forensic odontology identification

Sher-Lin Chiam [a,*], Denice Higgins [a], Kim Colyvas [b], Mark Page [c], Jane Taylor [c]

[a] Forensic Odontology Unit, The University of Adelaide, Australia
[b] School of Mathematical and Physical Sciences, University of Newcastle, Newcastle, Australia
[c] School of Health Sciences, University of Newcastle, Australia

ARTICLE INFO

Keywords:
Forensic odontology
Identification
INTERPOL Odontology scale
Judgment
Decision making
FORENSIC

ABSTRACT

Forensic odontology identification scales are used to express certainty of identifications of deceased persons. These standardized scales are assumed to convey unambiguous expert opinions and facilitate communication between forensic odontologists and end users. However, to date no studies have investigated how the experts interpret and use these scales.

Forensic odontology identification scales are used to express certainty of identifications of deceased persons. These standardized scales are assumed to convey unambiguous expert opinions and facilitate communication between forensic odontologists and end users. However, to date no studies have investigated how the experts interpret and use these scales.

This paper aims to examine the interpretation of the DVISYS forensic identification scale and choices of the levels in the scale subsequent to, and derived from, comparison of pairs of dental radiographs by extending the analysis of the data collected in the study by Page and Lain et. al. 2017.

The studied variables: self-reported confidence, forced binary decision of match and non-match, choice of level in the DVISYS scale (Identified, Probable, Possible, Insufficient and Exclude) were further analysed in this study using mixed models for relationships between the choices of level in the identification scale and the fundamental beliefs of likelihood of identification.

The results of this further analysis showed that the reported confidence of the decisions was correlated to the difficulty of cases, and as confidence decreased the use of less definitive terms ('Probable', 'Possible' and 'Insufficient') increased. 'Probable' and 'Possible' were used mainly in underlying beliefs below that of 'Identified' whereas 'Insufficient' was used mainly to convey a sublevel of 'Exclude'. The use of 'Insufficient' in this study was not consistent with the prescribed definition of the term.

The participants of the original study were not aware of the difficulty grading of the cases nor were required to grade them, however the reported confidence was systematically correlated to difficulty. Furthermore, indicated confidence level was correlated with choice of level on the scale in general, but the interpretation of the definition and application of the terms varied.

The findings reported here contribute to the foundational knowledge of factors governing the interpretation and application of the DVISYS forensic odontology identification scale and suggest that this scale may need to be modified.

1. Introduction

Forensic odontology identification helps to address whether a deceased person is the presumed identity by providing an opinion [1–3]. This opinion is presented as a degree of certainty, using a standardised categorical scale (e.g. ABFO, INTERPOL 2009 and INTERPOL 2013 -herein referred to as DVISYS scale), derived from comparison of antemortem and postmortem information and radiographs [4]. Prior to

* Corresponding author.
E-mail address: sherlin.chiam@gmail.com (S.-L. Chiam).

https://doi.org/10.1016/j.scijus.2021.02.007
Received 21 September 2020; Received in revised form 26 February 2021; Accepted 27 February 2021
Available online 9 March 2021
1355-0306/

reconciliation of the available data, dental evidence needs to be collected, collated, and interpreted. No published studies have to date validated the entire forensic dental identification process, however, studies that have been undertaken have predominately sought to validate the comparison of radiographs [5–14].

A study in 2017 by Page et al. [13], sought to examine the accuracy of dental radiograph comparison for identification by asking participants to provide a binary decision on whether a pair of radiographs belonged to the same or different individuals. Participants were also invited to use one or more of the three available standardised forensic odontology scales (ABFO, INTERPOL 2009, DVISYS) to provide an opinion on likely identity. The inclusion of the scales allowed participants to provide a categorical opinion primarily to simulate real world identification procedures. Whilst this study provided foundational validation of the method, it is important to build on this foundation and to investigate how the forensic odontology identification scales are used and interpreted.

The prescriptive nature of standardised terms and definitions used in identification scales is assumed to facilitate clear and uniform communication, and thus provide unambiguous interpretation. However, the extent to which the terms in the scale are uniformly interpreted and applied by odontologists has never been investigated. In the study by Page et al. [13], the term 'insufficient evidence', which is defined on the DVISYS scale as 'Neither PM nor AM comparison can be made' is surprisingly frequently associated with a correct binary decision with a high level of accuracy and confidence. This result may indicate that this category is not interpreted as defined but that some information was indeed available in these cases informing practitioner decision. If this is indeed the case, then this would lead one to question whether the definition of any of the categories on the scales are uniformly interpreted. Furthermore, although Page et al. [13] found a high inter-rater reliability in binary decision choices, substantial differences were noted in categorical scale selection within a case. A contributing factor to this may be variation in the degree of certainty reached, which is likely influenced by an individual's perception and rating of the difficulty of a case [15,16], despite these differing self-reported rating overall performance outcomes appeared to be systematically affected by the level of case difficulty as assigned by the study authors [17,18]. This performance and difficulty relationship was also observed in Page et al. [13]; accuracy rates were higher in easy cases compared to moderate and hard cases. Possibly, easy cases may be simple to decide, and this ease then may be reflected as high confidence decision. Research has shown that confidence corresponds to the cognitive ease of making a decision and hence by inference would be expected to correlated with difficulty [19,20].

Confidence and sense of difficulty may manifest in choice of level of identification following a binary decision. Anecdotally it is believed that the categorical levels in the scales express the practitioner's confidence of identification. However, in the study by Page et al. [13], correlation of reported confidence with choice of category on the scale was found to be inconsistent and non-linear as significant overlaps in the confidence ranges and thresholds for mid-level categories: 'Probable', 'Possible' and 'Insufficient' were reported. The assumption that the identification scale points are even simple representations of thresholds of confidence of identification was hence questioned by the authors of the study. Alternatively, the overlapping wide ranges may be the result of variable application of the mid-level terms on the scale because the underlying beliefs in match status are not specific. Other possible contributing factors could be personal bias in the confidence ranges; differing metacognitive or self-assessment of the confidence and accuracy relationship in individuals, and cognitive binning effect of categorisation (cut off points for each level in a scale) [21].

The results from original study [13] suggested that complex factors clearly governed the choice of the level of identification after the binary decision of match or non-match, and hence deserve further investigation. Variables recorded in the existing data: binary decisions, indicated confidence levels and identification level choices are reviewed in this current study to examine the relationships between the indicated confidence levels, the underlying belief of match status of the radiograph pairs and the choice of the level of identification. Additional analysis of this data may allow us to make inferences about the interpretation of and factors influencing the choice of identification category and to understand if the reported confidence level, level of difficulty and the believed match status were influencing factors in the decision-making process.

2. Materials and methods

The detailed research design and method for data collection can be found in the original paper [13] Therefore, only important aspects of the material and method are highlighted below.

- *Qualtrics*™ software platform [22] was used to present 50 paired antemortem and postmortem dental radiographs without any information provided. These were sampled from actual forensic odontology cases, where true match identification (n = 25) had been corroborated by DNA.
- The radiographs were graded by consensus by two practicing forensic odontologists into three levels of difficulty: 'Hard match' (n = 8), 'Moderate match' (n = 9), "Easy match" (n = 8), "Hard non match" (n = 7), 'Moderate non match' (n = 8) and 'Easy non match' (n = 10). Participants were not informed of the levels nor required to grade it.
- The participants viewed the dental radiograph image pairs, decided whether the images belonged to the same person, stated their confidence on a scale of 0–100% and finally chose a level of identification using one or all three of the recognised forensic odontology identification scales [ABFO [23]; INTERPOL scale (version 2009) [24] and DVISYS™ (DVI System International) which aligns with INTERPOL scale version 2013 [25] used in the PlassData Software)]. These standardised scales vary slightly between each other in levels of identification as shown in Table 1.

The responses of the forensic odontologists (n = 26) were extracted from this previous study and analysed to address the following research questions:

1. Was there a relationship between the level of difficulty (Easy, Moderate and Hard) and confidence level of the decisions?
2. Were there significant differences in confidence associated with the different levels in the DVISYS scale when analysed separately based on believed matched and non-matched status?
3. Did a participant report higher levels of confidence when the binary decision made was in fact correct? Was this level of confidence influenced by the perceived match and non-match status of the radiographic pair?
4. Did the level of difficulty influence the choice of level in the DIVSYS scale?
5. Were choices of "Probable', 'Possible' and 'Insufficient' used to convey underlying belief of match or non-match?
6. How did the participants align the different terms in the three identification scales when the scales consisted of slightly differing terms?

3. Analysis

Preliminary exploration was performed through box plots and bar charts, and subsequently mixed models were used to investigate significant effects due to non-independence and clustering of data: Linear and Generalised Linear Mixed Modelling (LMM, GLMM) [SPSS. Version 25 (SPSS, Inc; Chicago, Illinois) and SAS (SAS Institute Inc., V9.4 (TS1M4)]. Mixed models were chosen as the assumption of

Table 1
Terms used in the three forensic odontology identification systems.

Terms	DVISYS™/ INTERPOL version2013 [25]	ABFO [23]	INTERPOL version 2009 [24]
*Positive/ Identified	There is absolute certainty the PM and AM records are from the same person.	(Positive) The antemortem and postmortem data match in sufficient detail to establish that they are from the same individual. In addition, there are no irreconcilable discrepancies.	There is absolute certainty the PM and AM records are from the same person.
Probable	Specific characteristics correspond between PM and AM but either PM or AM data or both are minimal.	NA.	Specific characteristics correspond between PM and AM but either PM or AM data or both are minimal.
Possible	There is nothing that excludes the identity but either PM or AM data or both are minimal.	The antemortem and postmortem data have consistent features, but, due to the quality of either the postmortem remains or the antemortem evidence, it is not possible to positively establish dental identification.	There is nothing that excludes the identity but either PM or AM data or both are minimal.
Insufficient	Neither PM nor AM comparison can be made.	The available information is insufficient to form the basis for a conclusion.	NA.
Exclusion	Identity excluded (PM and AM records are from different persons).	The antemortem and postmortem data are clearly inconsistent.	PM and AM records are from different persons.

independence was most likely not met due to differences between matched pairs nested within participants and differences between participants. Therefore, two random intercepts were used in modelling, one for each factor.

Treating confidence as a continuous outcome, three linear mixed models (LMM) were fitted; firstly, with difficulty (Easy, Moderate or Hard) as the explanatory variable. The second model examined the DVISYS categories and perceived match status and their interaction. Thirdly, the relationship between accuracy (correct and incorrect decisions) and perceived match status (match or non-match) and their interaction were analysed. For all models, Restricted Maximum Likelihood (REML) estimation was used.

For the categorical outcome DVISYS the relationship with case difficulty and perceived match status was modelled with a multinomial distribution using a generalised logit link function and the DVISYS category 'insufficient information' as the reference category. Estimation failed with both factors in the model due to numerical problems because of one missing category in each believed match and non-match status (Exclude and Identified respectively). Therefore, separate models for each believed match status were run with case difficulty as the sole explanatory factor.

Pairwise comparisons using the fitted models were used to determine the significance of differences in levels of the explanatory factors, and statistical significance was set at 0.05. For the DVISYS outcome categories marginal probabilities and the 95% confidence intervals were determined for the difficulty levels using SAS.

To answer the question of how the participants, align and substitute the different terms used in the different scales, the pattern of usage in each scale was collated. The DVISYS scale was compared to INTERPOL and ABFO scale.

4. Results

Confidence ranges varied both between cases (Fig. 1a) and between participants (Fig. 1b) validating the need for random intercepts in the mixed models to account for these two sources of variation. This was confirmed by the significance of the effects when the models were run.

4.1. Was there a relationship between the different levels of difficulty (Easy, Moderate and Hard) and confidence levels of the decisions?

- Median confidence levels decreased with difficulty with wide and overlapping ranges for the moderate and hard cases [Easy ($n = 461$), Moderate ($n = 441$) and Hard ($n = 390$)]. (Fig. 2) As well, the variability of the distribution (range of the box and whisker section of the box plot) increased progressively from the Easy group with the smallest range to much wider for the Hard group. Notably in the Easy group where the confidence distribution was tightest there were a number responses that showed low confidence, being outliers ($\geq$1.5 times the IQR below the lower quartile, shown as circles) or extreme departures ($\geq$3.0 times the IQR, asterisks).
- This effect was shown to be significant using a Linear Mixed Model (LMM) with difficulty as the explanatory variable ($F(2,47) = 22.3$, $p < 0.001$).
- Marginal means and 95% confidence levels confirmed the increase in confidence with decreasing difficulty; hard $M = 63.2$; [56.1–70.2], moderate $M = 71.9$ [65.1–78.7], easy $M = 87.6$], [80.8–94.2].
- Pairwise comparison for confidence levels showed a significant difference between them all, $p < 0.001$.

In summary, the mean confidence level decreased significantly with increasing levels of difficulty.

4.2. Were there significant differences in confidence associated with the different levels in the DVISYS scale when analysed separately based on believed match and non-match status?

- Notably the definitive choices of 'Identified' and 'Exclude' were not ascribed respectively with non-match and match (match status as judged by the participants),

i.e., no-one indicated a case as identified when it was believed to be non-match and vice versa.

- Table 2 shows the frequencies of the DVISYS categories selected in believed matches and non-matches.
- A mixed model analysis with inclusion of interaction between the believed match status and DVISYS choice showed a significant interaction effect. $F(2,1230) = 17.35p < 0.001$.
- Table 3 summarises the mean marginal confidence and the 95% CI for the perceived matches and non-matches
- In matched decisions, the confidence levels decreased from 'Identified' to 'Insufficient' (most definitive to least definitive category). Pairwise comparison confirmed that all categories were significantly different ($p < 0.001$).
- In the perceived non-match group, the confidence decreased from the ascribed terms 'Exclude' to 'Possible' (definitive to less definitive choice), while confidence level for the ascribed term 'Insufficient' was comparable to 'Possible' (See Table 3). Pairwise comparison confirmed that only 'Exclude' was significantly different to other categories ($p < 0.001$).

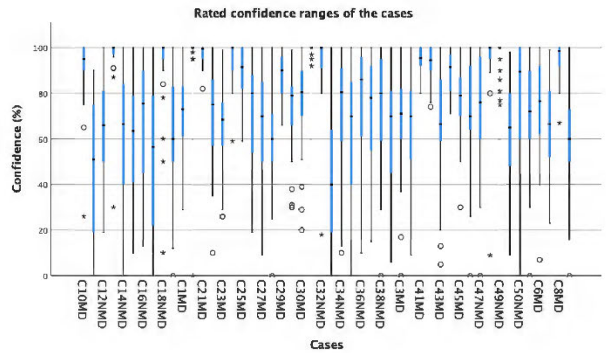


Fig. 1a. Box plots of rated confidence ranges of the cases.

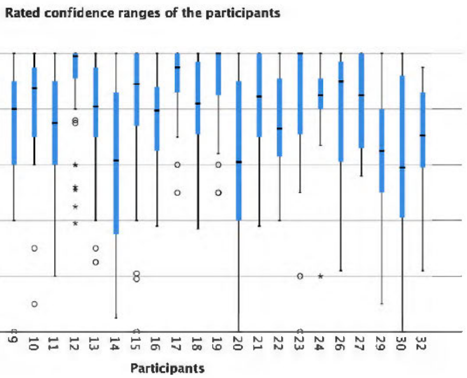


Fig. 1b. Box plots of rated confidence ranges of the participants.

- The mean confidence score for 'Probable' was significantly higher in the perceived match than non-match group, $p < 0.001$, but for 'Possible' it was not significantly different. For 'Insufficient' the confidence level was significantly lower in the perceived match than non-match group.

4.3. Did a participant report higher levels of confidence when the binary decision made was in fact correct? Was this level of confidence influenced by the perceived match and non-match status of the radiographic pair?

- Interaction effects for predictors; accuracy and believed match status were not significant ($F(1,75) = 2.61$, $p = 0.11$) for confidence in the LMM model. These main effects, however, were significant at $p <$ 0.001 and 0.01 respectively. Therefore, a main effects only model with fixed effects of believed match status and the accuracy was adopted, and this model was significant ($F(2, 1258)$ 60.8, $p < 0.001$).
- The predicted confidence levels in perceived matches were 75.2% and 59.0% for correct and incorrect decisions respectively, whereas in perceived non-matches it was 79.2% and 63% respectively (See Table 4). Overall, the confidence levels for correct decisions were 16% higher than incorrect decisions but incorrect non-matches were 4% higher than incorrect matches.

It appears that participants were more confident when their binary choice was indeed correct (match and non-match); more cautious with match decisions, and concomitantly less capable of assessing if they had

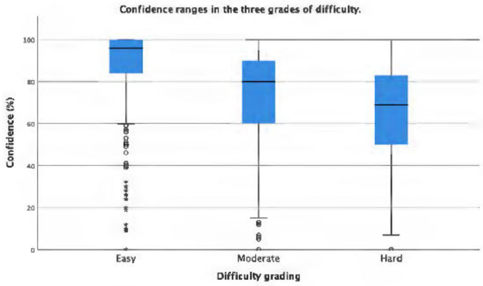


Fig. 2. Confidence ranges in the three grades of difficulty.

Table 2
Frequencies of DVISYS categories in each of the believed match status.

	DVISYS categories	Perceived Match	Perceived Non-Match	Total
	Identified	200	0	200
	Probable	200	7	207
	Possible	193	22	215
	Insufficient	88	247	335
	Exclude	0	335	335
Total		681	611	1292

Table 3
The marginal mean confidence and the confidence intervals for the DVISYS categories in perceived match and non-match.

Perceived	DVISYS	Mean	Std. Error	95% Confidence Interval	
				Lower	Upper
Perceived Match	Identified	92.8	2.2	88.3	97.2
	Probable	81.0	2.1	76.8	85.2
	Possible	60.6	2.1	56.4	64.7
	Insufficient	47.4	2.4	42.6	52.3
Perceived Non Match	Probable	60.0	6.0	48.3	71.8
	Possible	55.2	3.7	48.0	62.5
	Exclude	90.5	2.0	86.5	94.6
	Insufficient	57.1	2.1	53.0	61.3

Table 4
Fitted values and estimated means for confidence levels (%) for correct and incorrect decisions, cross tabulated with perceived match and non-match status of the radiographs.

	Correct Decision	Incorrect decision
Perceived match	75.2	59
Perceived non match	79.2	63

made correct non-match calls.

4.4. Did the difficulty level influence the probability of choice of the different levels in the DVISYS scale?

- Figs. 3a and 3b show the predicted marginal probabilities for the DVISYS categories across grades of difficulty from two GLMMs; one for each perceived match status. These models tested significant with F (6,471) = 5.72, $p < 0.001$) for believed match categories and (F (4,476) = 7.05,$p < 0.001$) for the non-match model.
- In believed non-matches (Fig. 3a) the likelihood of choosing 'Possible' was very low while categories 'Exclude' and 'Insufficient' predominated. The odds ratio (OR) of using 'Exclude' over 'Insufficient' was approximately 16 times higher 95% CI [5.4–49.5] in easy than in difficult cases but only two times higher in moderate compared to hard cases; this was marginally significant [0.7–6.0]. As difficulty increased, the likelihood of choosing 'Exclude' decreased while the likelihood of choosing 'Insufficient' increased.
- For perceived matches (Fig. 3b), there was a marked decrease in the use of 'Identified' from easy to hard cases. The OR of using 'identified' over 'Insufficient' was 72 times significantly higher in easy than in hard cases [12.7–414.3]. This decreased to approximately two times higher in moderate compared to hard cases, which was not significant [0.5–12.1]. However, the use of 'Probable' and 'Possible' increased in moderate and hard cases. In hard cases, the likelihood of using 'Possible' was the highest but was comparable to that for moderately difficult cases. The OR of choosing 'Probable' over 'Insufficient' was significantly seven times higher [2.3–18.8] for easy compared to hard cases, while it was nonsignificant and only twice [1.0–4.5] for moderate compared to hard. The ORs of using 'Possible' over insufficient were higher by two [1.0–5.5] and one and a half times [0.8–2.5] respectively in easy and moderate compared to hard cases.

In summary, definitive choices occurred most frequently in easy cases indicating a distinct difference between easy and hard cases which was not apparent between moderate and hard cases. In believed match cases of moderate and hard difficulty, there was a high probability of using 'Probable' and 'Possible' but when difficulty increased 'Insufficient' became more frequently used.

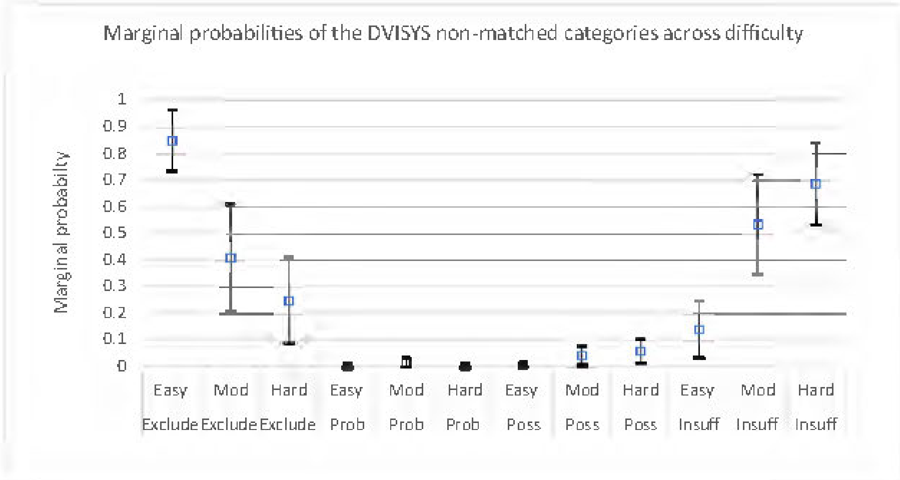
4.5. Were choices of "Probable', 'Possible' and 'Insufficient' used to convey beliefs of match or non-match?

- Fig. 4 summarises the marginal likelihoods of analysis of proportion of 'Probable' and 'Possible' and 'Insufficient' with perceived match status as the explanatory variable using the GLMM models. The ORs and 95% CI's for comparing 'Probable' and 'Possible' to 'Insufficient' were 60 [26–142] and 29 [17–49] times higher in the perceived match group than perceived non-match group respectively (both $p < 0.001$). The OR of choosing 'Probable' over 'Possible' was twice as high in the match than the non-match group. Notably, the OR of choosing 'Insufficient' over 'Possible' in the perceived non-match group was 27 [15–50] times higher than the perceived match group.

In summary, participants used 'Probable' and 'Possible' to convey beliefs at a sub level of identified match rather than non-match, whereas 'Insufficient' was used primarily when they believed it was likely a non-match.

4.6. How did the participants align the different terms in the three identification scales when the scales consist of slightly differing terms?

- A total of 22 participants used all three scales, one used INTERPOL and DVISYS while three used only the DVISYS scale. As the DVISYS scale comprised the full complements of categories, it was used as the basis for comparison with INTERPOL and ABFO scales because 'Insufficient' and 'Possible' were not available on these scales respectively.
- 'Identified' on DVISYS was aligned exclusively with the same term for both the ABFO (n = 175) and INTERPOL (n = 179). DVISYS 'Exclude' was aligned with on ABFO and INTERPOL 'Exclude' 97% (n = 279) and 99.7% (n = 300) of the time respectively.
- Where 'Probable' was available (DVISYS and INTERPOL scales) it was aligned exclusively (n = 175). On the ABFO scale, where probable was not available, 76% (n = 163) were reassigned as 'Possible'; the next level below identified, the remaining 24% were not realigned to any category on the ABFO scale.



Non-Match	Easy (CI)	Moderate (CI)	Difficulty (CI)
Possible	0.01 (0.00-0.01)	0.04 (0.00-0.08)	0.06 (0.02-0.11)
Insufficient	0.14 (0.03-0.24)	0.54 (0.34-0.73)	0.68 (0.52 -0.84)
Exclude	0.86 (0.75-0.97)	0.42 (0.21- 0.63)	0.26 (0.09-0.42)

Fig. 3a. Estimated Probabilities and 95% CIs of DVISYS for non-matched categories across grades of difficulty.

- In match decisions 'Possible' was aligned 97% and 100% of the time on the INTERPOL and ABFO scale respectively. The choice of 'Insufficient', available on DVISYS and ABFO scales, was realigned 99% of the time (n = 85) while it was replaced by 'Possible' 34% of the time on the INTERPOL scale (n = 86) with the remainder not realigned.
- In non-match decisions, 'Insufficient', present in both DVISYS and ABFO scale was aligned 97% of the time. (n = 206). On the INTERPOL scale where 'Insufficient' was absent, 63% were not aligned, 11% were aligned to 'Exclude' and 26% to 'Possible' (n = 212).

In summary definitive conclusions i.e., 'Identified' and 'Exclude' were aligned with the respective equivalent terms. For the sub-levels, the same term was often adopted when available on each scale despite differences in the hierarchical positions on the different scales. When the same term was not present, the nuance of the hierarchy and definition appeared to become important as seen in the high frequency of failure to align 'Insufficient' with another term on the INTERPOL scale.

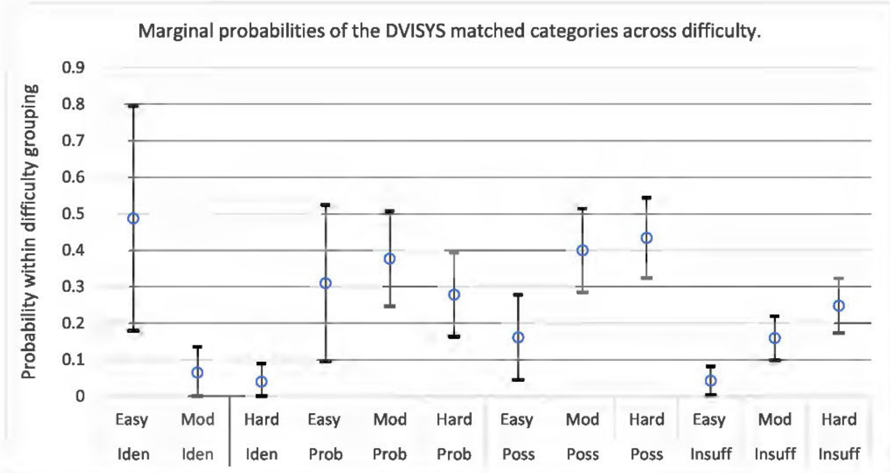
5. Discussion

This study aimed to investigate how the DVISYS forensic odontology identification scale was interpreted and applied by participants in a previous study by Page et al. [13] through mixed model analysis of the relationships of the recorded variables: binary decision, choice of level on DVISYS scale, perceived and actual match status, level of case difficulty (Easy, Moderate and Hard) and reported confidence levels.

High levels of confidence were seen in association with correct definitive decisions ('Identified' and 'Exclude') most frequently in easy cases. As 'Exclude' and 'Identified' where exclusively used in non-match and match decisions respectively this would suggest that the intended definitions for these two categories were clear and uniformly interpreted. While the forced binary decisions allowed assessment of fundamental accuracy results from this cannot be directly applied in practice as this is not how expert opinion is expressed in real casework. Nevertheless, apart from fundamental accuracy measure, the forced binary decisions used in this study allowed the verification of intended match status in the less definitive terms and hence were especially useful for elucidating the intent for use of the categories 'Probable' and 'Possible'.

The less definitive terms 'Probable' and 'Possible' were primarily associated with a binary choice of match, that is they were used to convey sub-levels of believed matches. The level 'Insufficient', however, was more often used with a binary non-match decision implying its interpretation as a sublevel of 'Exclude'. Confidence levels associated with 'Possible' and 'Insufficient' were not high (Table 3) suggesting that the terms may have been interpreted by the participants to mean 'inconclusive' or 'cannot be determined'. The apparent use of 'insufficient' evidence as a sublevel of 'Exclude' contradicts the definition of this term, which is "no comparison can be made". Interpretation of this term as a sublevel of 'Exclude' is further supported by the fact that participants, when selecting this category with a binary non-match decision, often did not align with any other level on the INTERPOL scale, which lacks an insufficient category and other existing terms were associated with match binary decisions. It is not possible to determine whether this is an aberration related to the forced binary decision or if individuals would have selected a sublevel of 'Exclude' if one were available.

As noted above, the scales are biased towards support for



Match	Easy (CI)	Moderate (CI)	Hard (CI)
Identified	0.49 (0.18-0.80)	0.06 (0.00-0.14)	0.04 (0.00-0.09)
Probable	0.31 (0.09-0.52)	0.38 (0.25- 0.51)	0.28 (0.16-0.39)
Possible	0.16 (0.04-0.28)	0.40 (0.28-0.51)	0.43 (0.32-0.54)
Insufficient	0.04 (0.00- 0.08)	0.16 (0.10-0.22)	0.25 (0.17-0.32)

Fig. 3b. Estimated Probabilities and 95% CIs of DVISYS for matched categories across grades of difficulty.

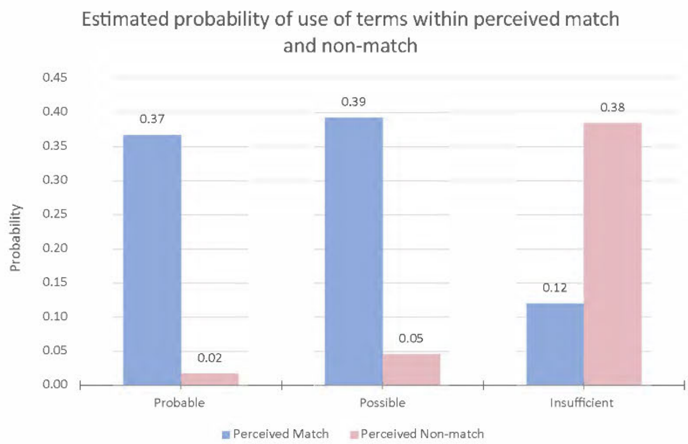


Fig. 4. Estimated likelihood of Probable, Possible and Insufficient in Perceived Match and Non-match.

identification with a lack of allowance for sub-levels of exclusion. It remains untested whether modification of a scale would enhance communication. For example, a scale with a different number of levels of identification, or explicit expressions of varying degrees of support for two opposing hypotheses. Uncertainty in how and if modification will be helpful is due to inherent shortcomings in all scales. Decision on choice of categories in a scale requires binning of continuous degrees of beliefs into thresholds and cut off points to fit into a scale - a phenomenon known as categorical perception [21,26,27]. This would inherently introduce individual cut-offs and threshold differences, although education has been shown to improve practitioners' calibration of the confidence threshold represented on the scale [28].

Confidence as measured in this study is not a simple construct. If the identification scale is assumed to represent confidence in the strength of the evidence, it should be directly correlated to reported confidence. However, the reported confidence rating is a very complex entity in decision making, due to bias in personal confidence ranges and participant differences in conceptual interpretation of confidence evaluation [29,30]. It could not be ascertained in this study if the confidence rating conveyed the strength of evidence or the metacognitive aspect of the choice. Strength of evidence pertains more to the likelihood of a match or non-match, whereas metacognition is about the awareness of the likelihood of correctness in the binary choice. Research suggests that both interpretations are possible [31]. Bias in reported confidence ranges and differing conceptual interpretation could possibly account for the lack of a clear linear relationship between confidence and DVI-SYS categories found in Page et al. [13]. These different interpretations and their association with confidence levels might be clarified by using a grading system that allows expression of the degree of certainty within the binary choices in future studies. In the original study by Page et al. [13] the confidence ranges overlapped for levels of 'Possible' and 'Insufficient'. One reason for this may be because these levels encompass both match and non-match decisions. In this current study, reanalysis of the data with separation of the match status demonstrated a more consistent relationship between confidence and the level of the scale chosen and thus supported this reasoning. The separation of believed match status in analysis of the confidence range made the relationship of the identification terms and confidence ranges more distinct.

Overall, incorrect non-match decisions were associated with higher confidence levels compared to incorrect believed matches. This effect might have been due to the small sample size or actual lower capability of self-assessment of accuracy (calibration) in believed non-matches than matches. Poorer calibration may be due to lack of practice, as most routine casework involves a likely identity indicated by circumstantial evidence. This is also supported by the lower accuracy rates for exclusions noted by both Pretty et al. [9] and Page et al. [13]. Pretty et al. [9] also posited that most routine cases are identifications rather than exclusions. Practice allows schemas and internal calibration for more efficient interpretation of evidence and information; confidence naturally increases with practice [32–34]. This might have contributed to a more consistent confidence relationship with the different terms in believed-match-cases.

Confidence appears to be a proxy for the perception of difficulty as levels of confidence decreased with increased difficulty. This sense of difficulty may not pertain exclusively to self-evaluation of accuracy; but also, to the difficulty of placement on the identification scale. Research has shown that confidence level is associated with cognitive processing fluency or 'cognitive ease' of making a decision which is facilitated by clarity and strength of available information [19,35,36]. Quality of radiographs, differences in the angulation, limitation of available similarities for comparison, commonly cited as determinants of difficulty [17,18,37], may translate to the level of cognitive difficulty and ease expressed as level of confidence. Inference of choice difficulty can be made from the higher confidence ranges in the choice of 'Insufficient' in believed non matches compared to believed matches. Lack of terms to accommodate non -match beliefs with lower confidence level made it difficult to choose 'Exclude', this might explain the paradoxically above chance accuracy rate of 78% in the original study when 'Insufficient' was defined as 'no information available for analysis', as noted above.

It can thus be inferred that formation of a final opinion and the associated reported confidence is a complex process even when the evaluation is purely on radiograph/image comparison. Hence, it is interesting to question if this comparison, evaluation, and decision-making process is influenced by the addition of information. Contextual information may induce a bias, anchor and influence the decision process especially in difficult cases [34].

6. Conclusion

The results of this study suggest that the scales used in decision making need to be reviewed or at the very least have more clearly defined meanings. Additionally, we show that confidence is affected by the level of case difficulty and relates to the difficulty in committing to a level on the identification scale. Difficult cases may be prone to the influence of contextual bias, confidence level could be used as a measure for alerting cases which will require non-case relevant information management. This study contributes to the foundational knowledge of factors influencing the choice and interpretation of the terms in the identification scale by forensic odontologists when evaluating radiographs for identification.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Taylor, J. Kieser (Eds.), Forensic Odontology: Principles and Practice, John Wiley & Sons, Incorporated, Chichester, West Sussex, UK; Hoboken NJ, 2016.
[2] J. Hinchliffe, Forensic Odontology, Part 1. Dental identification, Br. Dent. J. 210 (5) (2011) 219–224, https://doi.org/10.1038/sj.bdj.2011.146.
[3] I.A. Pretty, D. Sweet, A look at forensic dentistry — part 1: the role of teeth in the determination of human identity, Br. Dent. J. 190 (2001) 359–366.
[4] D. Higgins, H. James, Classifications used by Australian forensic odontologists in identification reports, J. Forensic Odontostomatol. 24 (2006) 32–35.
[5] K. Kaur Bhullar, R.S. Bhullar, S. Balagopal, A. Ganesh, M. Raian, Evaluation of dental expertise with intraoral periapical view radiographs for forensic identification, J. Forensic Dent. Sci. 6 (2014) 171–176, https://doi.org/10.4103/0975.
[6] G. Ekstrom, T. Johnsson, H. Borman, Accuracy among dentists experienced in forensic odontology in establishing identity, J. Forensic Odonlo-Stomatol. 11 (1993) 45–52.
[7] S. Fridell, J. Ahlqvist, The use of dental radiographs for identification of children with unrestored dentitions, J. Forensic Odontostomatol. 24 (2006) 42–46.
[8] S.L. Kogon, A review of validation studies of dental bitewing radiographs for forensic identification, J. Can. Soc. Forensic Sci. 29 (3) (1996) 113–117.
[9] I.A. Pretty, R.J. Pretty, B.R. Rothwell, D. Sweet, The reliability of digitized radiographs for dental identification: a web-based study, J. Forensic Sci. 48 (2003) 1–6.
[10] H. Borrman, H.-G. Grondahl, Accuracy in establishing identity by means of intraoral radiographs, J. Forensic Odonlo-Stomatol. 8 (1990) 31–35.
[11] A. Wenzel, A. Richards, J. Heidmann, Matching simulated antemortem and postmortem radiographs from human skulls by dental studens and experts: testing skills for pattern recognition, J. Forensic Odontostomatol. 28 (2010) 5–12.
[12] S.A. Sholl, G.H. Moody, Evaluation of dental radiographic identication: an experimental study, Forensic Sci. Int. 115 (2001) 165–169.
[13] M. Page, R. Lain, R. Kemp, J. Taylor, Validation studies in forensic odontology – part 1: accuracy of radiographic matching, Sci. Justice (2018), https://doi.org/10.1016/j.scijus.2017.11.001.
[14] S.-L. Chiam, M. Page, D. Higgins, J. Taylor, Validity of forensic odontology identification by comparison of conventional dental radiographs: a scoping review, Sci. Justice. 59 (2019) 93–101, https://doi.org/10.1016/j.scijus.2018.08.008.
[15] G. Eksrom, Accuracy among dentists experienced in forensic odontology in establishing identity, J. Forensic Odontostomatol. 11 (1993) 45–52.
[16] S.B. Balla, A. Forgie, Identification by comparison of caries free bitewing radiographs: impact of observer qualifications and their clinical experience, Forensic Sci. Criminol. 2 (2017) 1–5, https://doi.org/10.15761/FSC.1000108.
[17] H. Soomer, M.J. Lincoln, H. Ranta, A. Penttilä, E. Leibur, Dentists' qualifications affect the accuracy of radiographic identification, J. Forensic Sci. 48 (2003) 1121–1126.
[18] V. Pinchi, G.A. Norelli, F. Caputi, Dental identification by comparison of antemortem and postmortem dental radiographs: influence of operator qualifications and cognitive bias, Forensic Sci. Int. 222 (2012) 252–255.
[19] D. Fleisig, Adding information may increase overconfidence in accuracy of knowledge retrieval, Psychol. Rep. 108 (2011) 379–392, https://doi.org/10.2466/04.11.23.PR0.108.2.379-392.

[20] A.L. Alter, D.M. Oppenheimer, Uniting the tribes of fluency to form a metacognitive nation, Personal. Soc. Psychol. Rev. 13 (2009) 219–235, https://doi.org/10.1177/1088868309341564.

[21] I.E. Dror, G. Langenburg, "Cannot Decide": the fine line between appropriate inconclusive determinations versus unjustifiably deciding not to decide *, J. Forensic Sci. 64 (2019) 10–15, https://doi.org/10.1111/1556-4029.13854.

[22] Qualtrics software, Copyright ©, Qualtrics, Qualtrics and All Other Qualtrics Product or Service Names are Registered Trademarks or Trademarks of Qualtrics, Provo, UT, USA, 2016, http://www.qualtrics.com.

[23] ABFO, ABFO Body Identification Guidelines, (2000). http://abfo.org/wp-content/uploads/2012/08/ABFO-Body-ID-Information-Guidelines-Feb-2017.pdf (accessed June 18, 2017).

[24] INTERPOL Disaster Victim Identification Guide (Version 2009), (2009) 1–55. https://www.cmu.edu/chrs/conferences/eppi/docs/Interpol DVI Guide.pdf (accessed June 15, 2019).

[25] INTERPOL Disaster Victim Identification Guide (Version 2013), INTERPOL. (2014). DOI:10.1007/s13398-014-0173-7.2.

[26] S. Harnad, Categorical perception, in: Encycl. Cogn. Sci., Nature Publishing Group: Macmillan, 2003, https://doi.org/10.1002/0470018860.s00490.

[27] M. Fisher, F.C. Keil, The binary bias: a systematic distortion in the integration of information, Psychol. Sci. 29 (2018) 1846–1858, https://doi.org/10.1177/0956797618792256.

[28] K.A. Martire, B. Growns, D.J. Navarro, What do the experts know? Calibration, precision, and the wisdom of crowds among forensic handwriting experts, Psychon. Bull. Rev. 25 (2018) 2346–2355, https://doi.org/10.3758/s13423-018-1448-3.

[29] J. Navajas, C. Hindocha, H. Foda, M. Keramati, P.E. Latham, B. Bahrami, The idiosyncratic nature of confidence, Nat. Hum. Behav. 1 (2017) 810–818, https://doi.org/10.1038/s41562-017-0215-1.

[30] E. Norman, M.C. Price, Measuring consciousness with confidence ratings, in: Behav. Methods Conscious. Res., 2015, pp. 159–180, https://doi.org/10.1093/acprof:oso/9780199688890.003.0010.

[31] A. Pouget, J. Drugowitsch, A. Kepecs, C. Sciences, C.N. Unit, C. Spring, C. Spring, Confidence and certainty: distinct probabilistic quantities for different goals, Nat Neurosci. 19 (2016) 366–374, https://doi.org/10.1038/nn.4240.Confidence.

[32] I.E. Dror, D. Charlton, The paradox of human expertise: why experts get it wrong, in: N. Kapur (Ed.), Parad. Brain, Cambridge University Press, 2011, https://doi.org/10.1016/s1355-0306(98)72118-5.

[33] N. Charness, Tuffiash Michael, The role of expertise research and human factors in capturing, explaining, and producing superior performance Neil, Hum. Factors 50 (2008) 427–432, https://doi.org/10.1038/jid.2014.371.

[34] P.A. Fraser-Mackenzie, R.E. Bucht, I.E. Dror, Forensic Judgment and Decision Making, in: T.R. Zentall, P.H. Crowley (Eds.), Comp. Decis. Anal., Oxford University Press, 2013: pp. 385–415. DOI:10.1093/acprof:oso/9780199856800.003.0038.

[35] A. Koriat, S. Lichtenstein, B. Fischhoff, Reasons for confidence, J. Exp. Psychol. Learn. Mem. 6 (1980).

[36] P.D. Kvam, T.J. Pleskac, Strength and weight: the determinants of choice and confidence q, Cognition 152 (2016) 170–180, https://doi.org/10.1016/j.cognition.2016.04.008.

[37] D.F. MacLean, S.L. Kogon, L.W. Stitt, Validation of dental radiographs for human identification, J. Forensic Sci. 39 (1994) 1195–1200.

Chapter 5

Publication

Chiam S L, Dror IE, Huber CD, Higgins D. The biasing impact of irrelevant contextual information on forensic odontology radiograph matching decisions. Forensic Sci Int 2021;327:110997. https://doi.org/10.1016/j.forsciint.2021.110997.

Statement of Authorship

Title of Paper	The biasing impact of irrelevant contextual information on forensic odontology radiograph matching decisions.
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Chiam SL, Dror IE, Huber CD, Higgins D. The biasing impact of irrelevant contextual information on forensic odontology radiograph matching decisions. Forensic Sci Int. Elsevier; 2021; 327:110997. Available: https://doi.org/10.1016/j.forsciint.2021.110997

Principal Author

Name of Principal Author (Candidate)	Sher-Lin Chiam		
Contribution to the Paper	Conception and design of the project Acquiring of research data Analysis and interpretation of research Wrote manuscript Acted as corresponding author		
Overall percentage (%)	85%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	30th May 2022

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Itiel E. Dror		
Contribution to the Paper	Advised on concept and design Manuscript evaluation and editing		
Signature		Date	30th May 2022

Name of Co-Author	Christian D. Huber		
Contribution to the Paper	Data analysis Supervised and helped in data interpretation Manuscript evaluation and editing		
Signature		Date	31st May 2022

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Denice Higgins		
Contribution to the Paper	Supervised development of work, Helped in data interpretation Manuscript evaluation and editing		
Signature		Date	18th September 2022

Forensic Science International 327 (2021) 110997

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

The biasing impact of irrelevant contextual information on forensic odontology radiograph matching decisions

Sher-Lin Chiam[a,*], Itiel E. Dror[b], Christian D. Huber[c], Denice Higgins[a]

[a] Forensic Odontology Unit/Adelaide Dental School, The University of Adelaide, Australia
[b] Center for the Forensic Sciences, University College London, UK
[c] Department of Biology, The Pennsylvania State University, University Park, 16802 PA, USA

ARTICLE INFO

Article history:
Received 21 April 2021
Received in revised form 3 September 2021
Accepted 7 September 2021
Available online 10 September 2021

Keywords:
Forensic science
Forensic odontology
Identification
Cognitive bias
Contextual information

ABSTRACT

The potential biasing effect of irrelevant context information on the forensic odontology method of radiograph-based identification has never been empirically investigated despite being a recognized problem in other forensic science disciplines. This study examines the effect of irrelevant context information on the probability judgment of match (JOM) of practicing forensic odontologist and dentist participants who were asked to match pairs of dental radiographs supplemented with irrelevant case information.

The irrelevant case information contained domain task-irrelevant context information which varied in strength (strong or weak). It suggested either supportive or contradictory bias relative to the actual match status of the radiograph pairs. The dental radiographs consisted of verified match and non-match radiographs pairs sampled and de-identified from actual forensic cases. Changes in accuracy and JOM between supportive and contradictory contexts conditions revealed a contextual bias. Mixed model analysis showed that strong supportive context increased the odds ratio of correct decisions by a factor of 2.4 [1.23, 4.46]; $p = 0.0097$. Consistent with the biasing effect, the JOM score differences between strong supportive and contradictory irrelevant context information were 1.03 and 0.43 respectively for the non-match and match decisions. The direction of context suggestion ($p = 0.0067$), the radiograph match status ($p = 0.014$), and their interactions ($p = 0.0061$), were all found to impact the participants' decision. The weak context information was not strong enough to have a significant effect on accuracy or JOM scores.

This study demonstrates that radiograph match judgment is affected and can be biased by strong irrelevant contextual information.

1. Introduction

Cognitive bias, which pertains to systemic errors stemming from unconscious human perception and cognition when interacting with the external environment, affects judgment and decision, even by experts [1,2]. Cognitive bias has also been shown to specifically affect judgment and decision making in forensic science [1,3,4]. Bias was also identified as a concern that requires more research in an authoritative review undertaken by the National Academy of Science in the United States in 2009 [5]. Since this review, there has been increased recognition and research into this aspect of forensic science practice termed "cognitive forensics" [6].

There are multiple sources of cognitive bias, irrelevant contextual information is one of these [7,8]. In forensic science contextual task irrelevant information is considered as information that is not needed or relevant to the acquisition, analysis, comparison and evaluation of evidence for the specific expert opinion requested [7,9,10]. Irrelevant contextual information can be found in different areas and at different points in the forensic analysis process, some forms are imbedded in the evidence itself and are inseparable from the evidence, e.g., the content of a document examined by a handwriting expert, or the tone and emotional infliction in voice analysis. Another source is the circumstantial or police report that accompanies a case [6,8,10]. Exposure to irrelevant context information can give rise to cognitive bias impacting the expert's opinion, making their observation and conclusions no longer impartial [7,11].

Context effect influences the decision process as it may evoke certain expectations or hypotheses, causing a more top-down

* Corresponding author.
E-mail addresses: sherlin.chiam@gmail.com, sher-lin.chiam@adelaide.edu.au, sher-lin.chiam@adelaide.edu.au (S.-L. Chiam), i.dror@ucl.ac.uk (I.E. Dror), christian.domitian.huber@gmail.com (C.D. Huber), denice.higgins@adelaide.edu.au (D. Higgins).

https://doi.org/10.1016/j.forsciint.2021.110997
0379-0738/

approach to the evidence [12] This top-down approach can alter the attention, search, processing and perceived weight of the incoming information hence leading to cognitive bias [12]. Context effect can also shift the decision threshold; changing the quantity and quality of information required to commit to a decision [8]. This threshold has been shown to be very malleable to both internal and external factors. Internal factors include motivation, experience, and affective state of the decision maker, whilst external factors may include the time constraint, and the quality and quantity of available information [8].

Context effects can result in a change in the final outcome, i.e., the accuracy of the decision, or overall confidence in the decision. However, in some cases, the final outcome may not be an incorrect decision, but the threshold certitude of the final decision may change [8]. The magnitude of context effect is not only dependent on the strength and direction of the bias created by the circumstantial information but also on the quality of the actual evidence. Ambiguous or difficult evidence are more prone to context effect because multiple interpretations are possible; or the decision threshold is borderline between two categories [8].

Context effects have been shown to affect multiple disciplines including fingerprint analysis [11], blood splatter analysis [13], anthropology [14], mixed profile DNA evidence interpretation [15], and forensic pathology [16] because human decisions are required, and cognitive bias occurs at a subconscious level throughout the decision process. However, the level of susceptibility to context bias may vary for different disciplines [17]. Comparative or pattern matching forensic science disciplines that are more subjective are more prone to context bias because the majority of the evaluative and decision process depends on the discretion of the decision maker; nevertheless, bias impacts even objective forensic domains, such as toxicology [8,18]. However, this effect is more pronounced where there is a lack of databases and rule-based decision guidelines, or where the intermediate steps and rationale towards the conclusion are not clear [19].

Forensic human Identification is a comparative task requiring an estimate of the probability that the sample provided is derived from a known source. Identification by the process of dental matching requires comparison of antemortem and postmortem information. Information can include photographs, casts or scans of the dentition, treatment history and radiographs [20–23].

Unlike written clinical notes and charts, tangible forms of evidence such as radiographs, photographs, scans, and casts are less subjected to human error in the transcribing process hence are much more reliable forms of evidence. Orientation errors are possible with digital radiographs [24], nevertheless radiographs are still more reliable forms of evidence as such errors can be verified. Furthermore, radiographs allow visual verification and substantiation of the judged similarities and differences of the comparison [25]. However, comparison and evaluation of radiographs is not straightforward as radiographs taken of the same individual will differ when acquired on different occasions. Reasons for variation includes different angulation, exposure, contrast, and also biological, iatrogenic changes [26–32].

As there are no recognized data bases to allow development of a rule-based decision guideline for arriving at the probability of an identification for such comparisons, judgments and decisions depend on the tacit knowledge of the analyst [33,34]. Hence the judgment and decision process may be biased by contextual information.

Context information and resulting bias is a recognized problem, however, there has not been any empirical investigation of whether the forensic odontology method of identification is susceptible. There is a need to investigate the influence of contextual bias on identification by dental comparison. This paper aims to investigate the effect of different types of irrelevant context information on the accuracy and estimate of probability of match or non-match when comparing dental radiographs.

2. Method

2.1. Procedure

The study consisted of twelve experimental cases presented online using Qualtrics™ software [35]. For each case the participants had to initially read context information (circumstantial case information) prior to viewing and comparing a pair of dental radiographs. They then had to indicate how likely the pair of radiographs originated from the same person by using a 10-point scale ranging from −5 to +5 (−5 to −1 for indicating a non-match, and 1–5 for indicating a match; 0 was not an option). To provide a break and prevent carry over effect from one case to another, general knowledge questions were inserted between cases and had to be completed in order to progress to the next case. At the end of the experiment, participants were invited to answer questions about their professional experience, education, and history.

Before the experiment, an information statement and consent form was provided. In addition, participants had to watch an instructional video and complete a practice case before proceeding to the actual experimental trials. Participants were randomly assigned to one of two groups, both groups viewed the same radiograph pairs but were given contrasting context information associated to each radiograph pair.

2.2. Materials

The context information provided two independent variables with two levels each:

- strength of context (strong or weak) and
- direction of context (supportive or contradictory)

Strength was controlled through three main attributes:

- Consistency and coherence of the narrative,
- amount of definitive information [36] and
- placement of pertinent information in the narrative (beginning, middle or end) [37].

Supplementary method employed was modulating the certitude commitment in judgment or decisions by changing the cost of a wrong decision e.g., possible close scrutiny due to further investigations [2].

Direction of bias was achieved by suggesting an identification or an exclusion. Direction combined with strength provided four conditions: strong supportive, strong contradictory, weak supportive, and weak contradictory.

Main differentiating characteristics of strong from weak contexts was the greater consistency of supportive identifying or non-identifying facts. In addition, definitive statements about identity were placed at the start of narrative and, or at the conclusion (primacy and recency effects) [37]. Weak contexts by contrast included both supportive and contradictory facts. The primary direction of the context was determined by the quantity and quality of the supportive and contradictory facts [36]. Information about identity or non-identity was usually confined to the middle of the narrative.

In addition to the approach outlined above, subtle hints of the strength and direction of context were incorporated into the labeling at the start of each case before the participants progressed onto the actual scenario [38]. Strong exclusion vignettes contained only the case number as opposed to strong identification vignettes which displayed the name and birthdate of the deceased alongside the mortuary number. Weak exclusion vignettes were labeled with 'believed identity' whereas weak identification cases were labeled as 'presumed identity'. These more subtle differences were used to prime the intended manipulation of the strength and direction of context [38]. Defining elements were incorporated into each vignette to strengthen the intended manipulation. These elements provided four types of vignettes.

1. Strong identification: Definitive statements of the deceased being the presumed identity appeared in the first or second paragraph of the narrative. Consistent and coherent supportive information of identification. Concluding task statement directly linked the presumed identity to the deceased creating the final impression before radiograph evaluation.
2. Weak identification: These narratives lacked the definitive link of the identification at the beginning, implied or direct mention of presumed identity was incorporated into the middle of the narrative. They also lacked the direct identification documents or items on the deceased. Where there was direct mention of the deceased as the presumed identification, the narrative would have elements of other possible identification or circumstances that evoke higher cost of the conclusion if a mistake was made. These elements reduced the strength by virtue of increasing the weight and consequence of the opinion. Strength of suggestion was also weakened by incomplete information, e.g., other possible identification but records not yet available, or two bodies but only one record currently located.
3. Strong non-identification: The consistency and coherence in exclusion was maintained throughout the narrative in addition to reinforcing the possible wrong identification in the concluding statement. Placement of a statement implying previous wrong identification at the start e.g., "This is a request for assistance with clarification of identity". Wrong identification or non-identification was exacted through previous mistakes by other practitioners or strong evidence that the supposed deceased was still alive. Association of the presumed identity was minimized or absent even at the introduction of the case prior to scenario presentation. For example, only a postmortem number was presented, as opposed to strong identified where the mortuary number and name of the deceased were presented.
4. Weak Non-identification: Statements implying previous wrong identification were placed in the middle of the narrative rather than at the start. Documentation or proof of identity was not located with the deceased. Inclusion about doubts of identity coexist with presumed identity, creating a lack of consistency and coherence. Final paragraph contained no link of presumed identity to the deceased; only the mortuary number together with doubts of presumed identity placed just before the concluding task statement.

In addition to the four types of vignettes used in the actual experimental cases, non-test vignettes were used in the video demonstration, and in the practice case, as well as between blocks of test vignettes. These non-test vignettes functioned to induce the participants to read the test vignettes by reducing the possibility of the participants learning that the test vignettes contained contextual irrelevant information [39]. Non test vignettes contained domain relevant information: the dates of exposure of the antemortem and postmortem radiographs, which helps in deciding if the differences between the pair of radiographs are within the possible tolerance given the time differences.

The experimental vignettes were pilot tested on 10 general dentists, they were told to rate the probability of identification based on the narrative alone without radiographs. This allowed verification of the strength and direction. The vignettes were then refined and further edited to balance the length and effect of the narratives [40], after which the full experiment was pilot tested on two odontologists.

The vignettes were presented before exposure of the radiograph images. Each vignette was presented as progressive paragraphs to completion by the participants clicking on the screen. This encouraged contact with the vignettes, and completing the scenario was a compulsory step before the radiographs were presented.

The experimental cases were assigned to one of three blocks, so each block consisted of four vignettes, one of each type. The order of presentation of these four cases were randomized within each block but the order for three blocks was fixed. Between each block, the non-test vignettes were inserted. The vignettes were also randomly distributed between the two groups of participants; both groups assessed the same radiographs but saw either of the vignettes pair (supportive or contradictory) for the same radiographs.

2.3. Stimuli

The 12 radiographs pairs were taken from actual forensic odontology cases where the matched status was also verified by DNA. The radiographs in this study were a subset of radiographs originally used in another study to validate matching accuracy of radiographs in the absence of any information [41]. The match decision confidence and accuracy scores were 70% or lower and thus represented more difficult radiograph pairs. Each of the radiograph pairs consisted of pairing of either an orthopantogram with an intra oral radiograph or two intra oral radiographs. Balanced design was ensured by equal numbers of match and non-matching dental radiographs. Six of the twelve pairs were randomly assigned to each context strength i.e., 3 matched and 3 non-matched pairs for strong context group and similarly for the weak contexts. Within each context strength group, the same radiograph pair was presented with contrasting context (supportive and contradictory).

The dependent variable was the participants' judgment of match or non-match (JOM). The scores range was from −5 to +5: +1 to +5 for the strength of a match and −1 to −5 for the strength of a non-match (the midpoint of zero was not an option). Participants indicated their decisions by using this ten-point sliding scale with negative 5 and positive 5 indicating highest certainty of non-match and match respectively. The Likert like scale allowed quantification of the finer judgment process, whereby differences in mean score between supportive (positive) and contradictory contexts (negative) suggests judgment difference for the same pair of radiographs. JOM scores also allowed assessment of accuracy, correct versus incorrect decisions, a more reductive binary measure, to be computed: negative scores are correct for non-matching radiograph pairs and positive scores are correct for matching pairs.

2.4. Participants

77 participants were recruited via international professional societies, 21 were excluded because of failure to complete the experimental task. The remainder consisted of 24 practicing dentists and 32 forensic odontologists (N = 56) from various countries (Australia, Finland, India, Italy, Norway, South Africa, Sweden, and the United States). The main criteria for inclusion was that they were currently practicing in their professions. Four forensic odontologists and two dentists did not fully complete the work experience questionnaire but identified their profession in the open response section. 42% of the dentists (n = 24) had more than 20 years of general dentistry practice experience (range: less than 5 to more than 20 years). Three had graduate diploma or higher in forensic odontology, six attended non-award short courses as part of continuing professional development however none of the dentists practiced forensic odontology. 28% of the forensic odontologists (n = 32) had more than 20 years' experience (range: less than 5 to more than 20 years). Only

two in this group did not have dental background; one was a forensic medical practitioner and another had a forensic science background. While four participants did not respond to questions about forensic odontology education; 8% did not undergo any course, 17% had been through non award courses and 75% had a graduate diploma or higher. 41% were registered as specialists in forensic odontology.

3. Results

Descriptive analysis provided an overview of the data and inferential statistics were then applied for further analysis. The data was clustered and non-independent as the same radiograph pairs were judged and rated by all participants. The rating would be more similar for the same participant and same radiograph [7], consequently Linear and Logistic Mixed models [R, Version 4.0.3 and function lmer and glmer of the package lme4] were used to analyze the data. Estimated marginal means were computed with the R package ggeffects.

Judgment of match (JOM): Mean JOM score varied between the supportive and contradictory, reflecting that decisions were affected by the irrelevant contextual information. The differences in mean JOM scores between the contradictory and supportive context was greater in the strong than the weak context group (see Table 1). The greatest mean difference was in the strong context group for non-matching radiograph pairs (1.00), while the lowest was in the weak context non-matched pairs (−0.08). Of note, all the mean and median JOM scores were aligned with the actual match status of the radiographs; matched pairs were positive and non-match pairs were negative.

Inferential statistical were then computed to examine significance using a generalized mixed model with JOM score fitted with these independent variables: strength of context, direction of context, radiograph match status and participant type, including an interaction between the strength and direction of context. The results showed that radiograph match status was significant. $p < 0.001$. Therefore, it appears that the match status of the radiographs had the strongest effect on the scores. However, as strong context appeared to affect the JOM scores, the strong and weak data were further analyzed by separate models.

The JOM was regressed fitting the variables: direction of context and radiograph match status and their interaction. Results showed that the direction of context, radiograph match status and their interactions were all significant. $p = 0.0067$; $p = 0.014$; $p = 0.0061$ (respectively).

The significant interaction effect suggested different effects of direction on the JOM scores for the different match status; the marginal means estimates provided further insights into this interaction. Fig. 1a illustrates the interaction effect; the mean JOM score difference between supportive and contradictory context was 1.03 and 0.43 for the non-match and match respectively. Furthermore, the JOM absolute value was larger (higher positive and negative) with supportive compared to contradictory context especially for the non-match decisions. Therefore, strong supportive context increased the JOM scores overall but was more pronounced in non-match than in match decisions compared to strong contradictory context. $p = 0.01$. Weak data group condition modeling showed that only the match status had a significant effect on the outcome. Interaction effect from the context direction and radiograph match status was not significant. Furthermore, the predicted mean JOM scores were comparable in both matching and non-matching decisions between supportive and contradicting context (see Fig. 1b). Therefore, weak context did not have any significant effect on the JOM score.

In conclusion, although the overall mean JOM scores were in accordance with the true match status of the radiographs, judgments were still affected by irrelevant contextual information. In particularly, the strong supportive context increased the certainty in a decision, especially in non-match decisions. Conversely weak context had no significant effects on either match or non-match decisions.

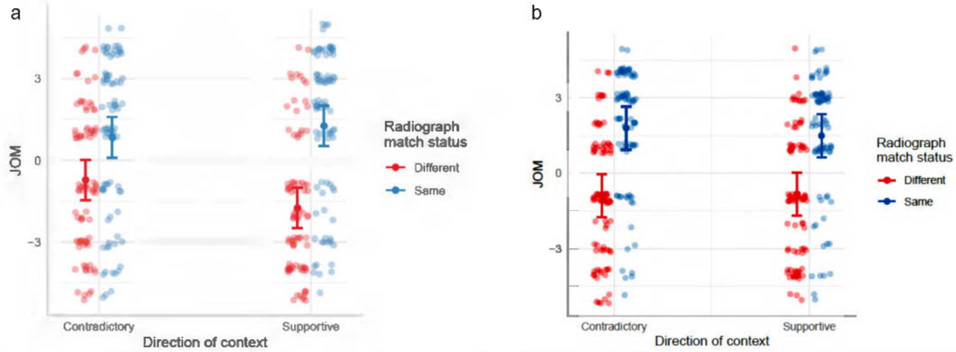
Accuracy: Judgment threshold (JOM) changes due to context effect may cumulate in alteration in accuracy (decision outcomes), hence accuracy was also computed. Accuracy scores or the frequency of correct decisions in each of the context conditions were obtained by condensing the judgment of match scores (JOM) into binary scores (See method section).

Descriptive data showed that strong supportive context increased the number of correct answers in both match and non-match decisions. The percentage increase of correct decisions was 7% and 18.7% for match and non-match decisions respectively, whereas in the weak context group, they were 1.6% and −2.5% (Table 2). Thus, in accordance with the analysis of JOM, strong supportive context increased accuracy over the contradictory context, whereas weak context did not seem to have any substantial effect on accuracy of decision.

Statistical significance and estimated odds were assessed using correct decisions as a binary outcome in a logistic mixed-effects model fitted with the independent variables of strength, direction of context, radiograph match status, and participant groups. An interaction effect between the strength and the direction of bias was also

Table 1
Judgment score (JOM) in the strong and weak irrelevant contextual conditions for the different radiograph pairs.

Strong context				
	Different person X-ray (non-match)		**Same person X-ray (match)**	
	Contradictory (N = 84)	**Supportive (N = 87)**	**Contradictory (N = 84)**	**Supportive (N = 87)**
Judgment score (JOM) (−5 to −1 and 1–5)				
Mean (SD)	−0.74 (2.53)	−1.74 (2.51)	0.81 (2.85)	1.30 (2.79)
Median [Min, Max]	−1.00 [−5.00, 4.00]	−2.00 [−5.00, 4.00]	1.00 [−5.00, 5.00]	2.00 [−5.00, 5.00]
Difference in mean between supportive and contradictory context	1.00		0.49	
Weak context				
	Different person X-ray (non-match)		**Same person X-ray (match)**	
	Contradictory (N = 87)	**Supportive (N = 84)**	**Contradictory (N = 87)**	**Supportive (N = 84)**
Judgment score (JOM) (−5 to −1 and 1–5)				
Mean (SD)	−0.90 (2.45)	−0.82 (2.56)	1.80 (2.40)	1.49 (2.45)
Median [Min, Max]	−1.00 [−5.00, 4.00]	−1.00 [−5.00, 5.00]	3.00 [−5.00, 5.00]	2.00 [−5.00, 5.00]
Difference in mean between supportive and contradictory	−0.08		−0.31	



	Strong context	Predicted scores	95% CI
Different person X-ray (non-match)	Contradictory	-0.72	[-1.46, 0.02]
	Supportive	-1.75	[-2.49, -1.02]
	Score difference between contexts	1.03	
Same person X-ray (match)	Contradictory context	0.84	[0.10, 1.59]
	Supportive context	1.27	[0.53, 2.00]
	Score difference between contexts	0.43	

	Weak context	Predicted scores	95% CI
Different person X-ray (non-match)	Contradictory	-0.89	[-1.75, -0.03]
	Supportive	-0.83	[-1.69, 0.04]
	Score difference between contexts	0.06	
Same person X-ray (match)	Contradictory context	1.80	[0.94,2.66]
	Supportive context	1.50	[0.63,2.36]
	Score difference between contexts	0.30	

Fig. 1. a. Predicted JOM (judgment of match) scores for match and non-match radiographs in the presence of strong supportive and contradictory context. b. Predicted JOM (judgment of match) scores for match and non-match radiographs in the presence of weak supportive and contradictory context.

included as suggested by the descriptive analysis. Results show the direction of context was significant p = 0.013, supportive context increased the odds of making a correct judgment by a factor of 1.81, 95% CI [1.14, 2.87] compared to contradictory context. The interaction effect of the strength and direction of context was marginally significant at p = 0.077 (Fig. 2).

As an interaction effect of strength was suggested in the descriptive analysis and modeling, separate mixed models were computed for the strong and weak data fitted with: direction of context and radiograph match status with the inclusion of their interaction only.

The strong context model showed that only the direction of context was significant at p = 0.0097, and the odds of a correct decision increased by a factor of 2.4 [1.23, 4.46] for supportive compared to contradictory context. Predicted marginal probabilities from this model provided further insights into the results (see Fig. 3a). Both matched and non-matched accuracy probability increased with strong supportive compared to contradictory context as mentioned above. Probability difference between supportive and contradictory context was 19% and 7% for non-match and match decisions respectively, i.e., the difference was greater in non-match decisions. However, overall, the interaction effect was not significant, therefore context did not have a different effect on the probability of making correct match and non-matched decisions.

For the weak context group, the differences in the predicted marginal probabilities were not remarkably different, (Fig. 3b) which concurred with the JOM analysis. Changes in the decision accuracy between the supportive and contrasting context was not observed

Table 2
Accuracy under the influence of strong and weak irrelevant contextual information for the different types of radiographs.

Strong context	**Different person X-ray (non-match)**		**Same person X-ray (match)**	
	Contradictory (N = 84)	**Supportive (N = 87)**	**Contradictory (N = 84)**	**Supportive (N = 87)**
Accuracy				
Incorrect answer	35 (41.7%)	20 (23.0%)	30 (35.7%)	25 (28.7%)
Correct answer	49 (58.3%)	67 (77.0%)	54 (64.3%)	62 (71.3%)
Difference between the supportive and contradictory for correct answers %	18.7		7	
Weak context	**Different person X-ray (non-match)**		**Same person X-ray (match)**	
	Contradictory (N = 87)	**Supportive (N = 84)**	**Contradictory (N = 87)**	**Supportive (N = 84)**
Accuracy				
Incorrect answer	32 (36.8%)	33 (39.3%)	19 (21.8%)	17 (20.2%)
Correct answer	55 (63.2%)	51 (60.7%)	68 (78.2%)	67 (79.8%)
Difference between the supportive and contradictory for correct answers%	−2.5		1.6	

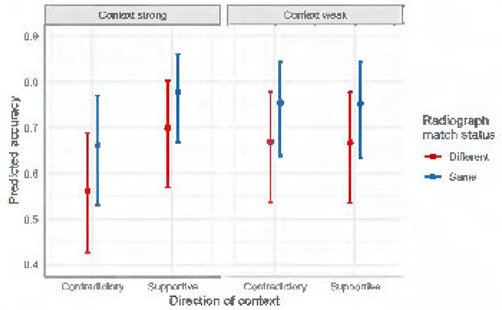


Fig. 2. Predicted probabilities of accuracy with predictors: radiograph match status, strength and direction of context.

for match nor non-match decisions. The most significant factor in predicting correct decisions is strong context direction alone. Weak context by contrast did not have any significant effects on accuracy.

4. Discussion

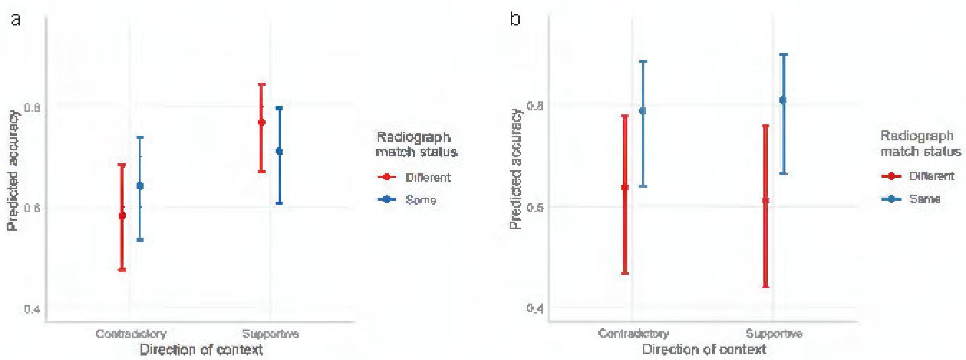
This experimental study with practicing professionals sought to investigate the effect of irrelevant contextual information on forensic radiographic matching decision, which is one of the key activities in human identification. Strong contextual information was shown to affect the judgment process as demonstrated by the change in certitude and outcome between supportive and contrasting context.

The outcome or the accuracy of the decision in this study increased with strong supportive context, with strong supportive context increasing the odds of making both correct match and non-match decisions. However, it is important to be cognizant of the fact that it cannot be argued that the supportive context made the decisions more accurate. These decision changes were made under the influence of irrelevant contextual information, and thus were not impartial and not based only on the evidence.

Accuracy signifies the crossing of a categorical decision threshold which in a real-world situation can have serious legal and ethical implications. Decision threshold or the certitude of match probability judgment of the radiograph pairs (Judgment of match-JOM scores) provided a window into the finer cognitive process and judgment threshold. The changes in the JOM scores between supportive and contrasting context provided a proxy measure of the context effect on the cognitive process and threshold of the matching process.

Decision threshold change is important and must be given consideration when examining context effect even when a correct decision is made [42]. Context effects can be dynamic and accumulative see bias cascade and bias snowball effects [43]. Consequently, it is difficult to predict if and when the effect is strong enough to alter the decision threshold significantly enough to alter the outcome. Furthermore, context effect is not restricted to written reports but also the professional social and other collaborative interaction with various stakeholders inherent within the practice environment (see Fig. 1 in [44]) which can affect judgments and decisions of the practitioners [8,10,45–47]. Another contribution to context effect is the quality and clarity of the evidence itself, which may also have influenced the decision-making process [8].

Evidence that is ambiguous or difficult to decide is more susceptible to the influence of context effect, because ambiguity and difficulty imply that the decision threshold is ambivalent between two boundaries and any additional information can potentially tip that balance over to either category [10,48]. Ambiguity and difficulty factors existed in the radiographs: ambiguity was due to the absence of relevant information e.g., exposure dates and treatment history. This information supports reconciliation of discrepancies between antemortem and postmortem radiographs, and the absence of this information makes resolving and explaining observable differences more equivocal leaving room and freedom for effect of context. This is especially relevant in this study as the radiograph pairs utilized



	Probability of correct identification and 95%CI	Probability of correct exclusion and 95%CI
Strong supportive	0.71 [0.61,0.80]	0.77[0.67, 0.85]
Strong Contradictory	0.64 [0.54,0.74]	0.58[0.48, 0.68]

	Probability of correct identification and 95%CI	Probability of correct exclusion and 95%CI
Weak supportive	0.81 [0.67, 0.90]	0.61[0.44, 0.76]
Weak contradictory	0.79 [0.64, 0.89]	0.64[0.47, 0.78]

Fig. 3. a. Predicted probability of correct decisions for match and non-match radiographs in the presence of strong supportive and contradictory radiographs. b. Predicted probability of correct decisions for match and non-match radiographs in the presence of weak supportive and contradictory radiographs.

were selected from ones classified as more difficult cases implying match categorizing difficulties. In this study we noted that non-match decisions were made with less certainty and accuracy than match decisions, as has been seen in previous research [41,49].

Unlike during a Disaster Victim Identification event a decision of exclusion is rare in routine forensic odontology practice as there is usually only one "believed to be" identity [22,50]. The lack of practice at excluding identity thwarts the development of schema and confidence in these decisions. A lower level of confidence in the decision outcome could potentially induce more reliance on the context information to aid decision [36]. This may account for the more pronounced context effect in non-match decisions as seen in the results; there was higher certainty and accuracy with strong supportive context in non-match decisions compared to match decisions. Conversely, weak context information did not have an observable effect on match or non-match decisions, which may suggest that matching radiographs is relatively robust and is only affected by strong context. Alternatively, the difference noted between strong and weak context seen here may be a result of the challenges seen in the designing of the vignettes.

Effective vignettes needed to be believable and relatable to actual real-life situations [51]. They also required distinguishable differences in the narratives to evoke the different levels of direction and strength [52]. The actual purpose of the vignettes could not be overt or obvious to the participants for the vignettes to exert the intended effect because context effect is subliminal and unconscious. Finally, the vignettes must also be read by the participants, therefore a balance of these factors was a major consideration for the design of the context information.

Strong non-match vignettes had the most direct and obvious suggestions of the four classes of vignettes. Non-match circumstantial evidence is very uncommon, the most probable situations where this might happen are revisits of prior incorrect identification. Paradoxically, this allowed imperative task sentences to be incorporated at the start and end of the narratives without reducing credibility whilst allowing for strong first and last impression of the direction and strength of the context. Strong match vignettes on the other hand were modeled on written police reports, this meant applying the "base-rate" effect because routine cases are mostly confirmation of a presumed identification. In strong match vignettes, the use of direct semantic association of the deceased with presumed identity, together with the consistency and coherence of the narrative increased the strength of the suggestion. It was reasoned that the same imperative sentences at the start would have reduced credibility and that it would have alerted participants to the objective of the study, unlike for strong non-match narratives. Weak vignettes by contrast required inclusion of both "push and pull" factors to control for strength. The coexistence of match and non-match information, which were the pertinent content, were placed in the middle of the narrative. This could have made it confusing and hence, may have made it difficult for this information to be assimilated and applied during the complex task of evaluating and comparing the radiographs and thus may have accounted for the lack of context effects.

In this study, there was no significant difference seen between dentists and forensic odontologists, this could be due to the small sample size or alternatively, due to the common skill set, basic dental training, between the two groups. The definition of forensic odontologist is varied across geographical regions, forensic odontologists in this study were self-identified as practicing forensic odontologists in an attempt to represent the real world sample. The dentists in this group did not practice forensic odontology although a few had some form of forensic odontology education. Furthermore, notably, almost half of the sample had considerable general dental practice experience (more than 20 years). Dental practice experience may have a more significant impact than formal forensic odontology education. Dental training would presumably have provided the basic perceptual ability for interpreting and evaluating radiographs, and matching the radiographs was effectively a perceptual task in this study as no relevant information was attached. The effect of expertize in forensic odontology may not be apparent and discernible in match or non-match decisions but rather in the determination of the probability of identification as represented by the choice on the INTERPOL forensic odontology identification scale. Determination of identification, probable, possible, exclusion and insufficient information is conceptually different from matching [53,54], although there is an assumption that the implicit match judgment precedes the decision of the final opinion. An opinion on identification requires an assessment of the weight of the evidence in light of all the relevant information and also the consequence of the opinion [55,56]. Expertize in forensic odontology rests on this ability to assimilate the tacit knowledge in dentistry, forensic science and the legal aspect of evidence. For future research into the effect of context information to be more applicable, the identification decisions in addition to match and non-match judgments should be investigated.

5. Conclusion

This exploratory study showed that in a controlled experimental environment, strong context information had an effect on radiograph matching judgments, especially in making non-match judgments. Although the majority of decisions were correct, judgment certainty and accuracy were affected by strong context.

CRediT authorship contribution statement

Sher-Lin Chiam: Conceptualization, Methodology, Investigation, Writing – original draft, Data curation. **Itiel Dror:** Methodology, Writing – review & editing. **Christian D. Huber:** Formal analysis, Writing – review & editing. **Denice Higgins:** Supervision, Methodology, Writing – review & editing.

Acknowledgments

We thank the two anonymous reviewers for their comments.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.forsciint.2021.110997.

References

[1] D. Kahneman, O. Sibony, C. Sunstein, Noise: A Flaw in Human Judgment, William Collins Publishers, Glasgow, Scotland, 2021.

[2] R.S. Nickerson, Confirmation bias: a ubiquitous phenomenon in many guises, Rev. Gen. Psychol. 2 (1998) 175–220, https://doi.org/10.1037//1089-2680.2.2.175

[3] J. Kukucka, I.E. Dror, Human factors in forensic science: psychological causes of bias and error, in: D. DeMatteo, K.C. Scherr (Eds.), The Oxford Handbook of Psychology and Law, Oxford University Press, New York, 2021(in press).

[4] G.S. Cooper, V. Meterko, Cognitive bias research in forensic science: a systematic review, Forensic Sci. Int. 297 (2019) 35–46, https://doi.org/10.1016/j.forsciint.2019.01.016

[5] Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council, Strengthening Forensic Science in the United States: A Path Forward, National Academy of Sciences, Washington, DC, 2009.

[6] I.E. Dror, Cognitive neuroscience in forensic science: understanding and utilizing the human element, Phil. Trans. R. Soc. B. 370 (2015) 1–8, https://doi.org/10.1098/rstb.2014.0255

[7] I.E. Dror, A hierarchy of expert performance, J. Appl. Res. Mem. Cogn. 5 (2016) 121–127, https://doi.org/10.1016/j.jarmac.2016.03.001

[8] I.E. Dror, Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias, Anal. Chem. 92 (2020) 7998–8004, https://doi.org/10.1021/acs.analchem.0c00704

[9] B.O. Gardner, S. Kelley, D.C. Murrie, I.E. Dror, What do forensic analysts consider relevant to their decision making? Sci. Justice 59 (2019) 516–523, https://doi.org/10.1016/j.scijus.2019.04.005

[10] G. Edmond, J.M. Tangen, R.A. Searston, I.E. Dror, Contextual bias and cross-contamination in the forensic sciences: The corrosive implications for investigations, plea bargains, trials and appeals, Law Probab. Risk 14 (2014) 1–25, https://doi.org/10.1093/lpr/mgu018

[11] I.E. Dror, D. Charlton, A.E. Péron, Contextual information renders experts vulnerable to making erroneous identifications, Forensic Sci. Int. 156 (2006) 74–78, https://doi.org/10.1016/j.forsciint.2005.10.017

[12] C.L. Maclean, I. Dror, A primer on the psychology of cognitive bias, in: C.T. Robertson, A.S. Kesselheim (Eds.), Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science, and Law, first ed., Academic Press, 2016, pp. 13–24.

[13] N.K.P. Osborne, R. Zajac, M.C. Taylor, Bloodstain pattern analysis and contextual bias, in: A. Jamieson, A. Moenssens (Eds.), Wiley Encyclopedia of Forensic Science, John Wiley & Sons, Ltd, Chichester, UK, 2015, pp. 1–7, https://doi.org/10.1002/9780470061589.fsa1127

[14] S. Nakhaeizadeh, I.E. Dror, R.M. Morgan, Cognitive bias in forensic anthropology: visual assessment of skeletal remains is susceptible to confirmation bias, Sci. Justice 54 (2014) 208–214, https://doi.org/10.1016/j.scijus.2013.11.003

[15] I.E. Dror, G. Hampikian, Subjectivity and bias in forensic DNA mixture interpretation, Sci. Justice 51 (2011) 204–208, https://doi.org/10.1016/j.scijus.2011.08.004

[16] I. Dror, J. Melinek, J.L. Arden, J. Kukucka, S. Hawkins, J. Carter, D.S. Atherton, Cognitive bias in forensic pathology decisions, J. Forensic Sci. 66 (2021) 1751–1757, https://doi.org/10.1111/1556-4029.14697

[17] I. Dror, The ambition to be scientific: human expert performance and objectivity, Sci. Justice 53 (2013) 81–82, https://doi.org/10.1016/j.scijus.2013.03.002

[18] H.J. Hamnett, I.E. Dror, The effect of contextual information on decision-making in forensic toxicology, Forensic Sci. Int. Synerg. 2 (2020) 339–348, https://doi.org/10.1016/j.fsisyn.2020.06.003

[19] President's Council of Advisors on Science and Technology, Report to the President Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, Washington DC, 2016.

[20] J. Taylor, J. Kieser (Eds.), Forensic Odontology: Principles and Practice, John Wiley & Sons, Incorporated, Chichester, West Sussex, UK; Hoboken, NJ, 2016.

[21] D. Sweet, Forensic dental identification, Forensic Sci. Int. 201 (2010) 3–4, https://doi.org/10.1016/j.forsciint.2010.02.030

[22] I.A. Pretty, D. Sweet, A look at forensic dentistry—part 1: the role of teeth in the determination of human identity, Br. Dent. J. 190 (2001) 359–366.

[23] J. Hinchliffe, Forensic odontology, part 1. Dental identification, Br. Dent. J. 210 (2011) 219–224, https://doi.org/10.1038/sj.bdj.2011.146

[24] S. Chiam, A note on digital dental radiography in forensic odontology, J. Forensic Dent. Sci. 6 (2014) 197–201, https://doi.org/10.4103/0975-1475.137072

[25] A. Forrest, Collection and recording of radiological information for forensic purposes, Aust. Dent. J. 57 (2012) 24–32, https://doi.org/10.1111/j.1834-7819.2011.01658.x

[26] K. Kaur Bhullar, R.S. Bhullar, S. Balagopal, A. Ganesh, M. Raian, Evaluation of dental expertise with intraoral periapical view radiographs for forensic identification, J. Forensic Dent. Sci. 6 (2014) 171–176, https://doi.org/10.4103/0975

[27] S.L. Kogon, A review of validation studies of dental bitewing radiographs for forensic identification, Can. Soc. Forensic Sci. J. 29 (1996) 113–117.

[28] D.F. MacLean, S.L. Kogon, L.W. Stitt, Validation of dental radiographs for human identification, J. Forensic Sci. 39 (1994) 1195–1200.

[29] S.A. Sholl, G.H. Moody, Evaluation of dental radiographic identication: an experimental study, Forensic Sci. Int. 115 (2001) 165–169.

[30] A. Wenzel, A. Richards, J. Heidmann, Matching simulated antemortem and postmortem radiographs from human skulls by dental studens and experts: testing skills for pattern recognition, J. Forensic Odontostomatol. 28 (2010) 5–12.

[31] S.B. Balla, A. Forgie, Identification by comparison of caries free bitewing radiographs: impact of observer qualifications and their clinical experience, Forensic Sci. Criminol. 2 (2017) 1–5, https://doi.org/10.15761/FSC.1000108

[32] H. Borrman, H.-G. Grondahl, Accuracy in establishing identity by means of intraoral radiographs, J. Forensic Odonlostomatol. 8 (1990) 31–35.

[33] C. de Villiers, V. Phillips, Person identification by means of a single unique dental features, J. Forensic Odontostomatol. 16 (1998) 17–21.

[34] A. Acharya, J.A. Taylor, Are a minimum number of concordant matches needed to establish identity in forensic odontology? J. Forensic Odontostomatol. 21 (2003) 6–13.

[35] Qualtrics software, Copyright ©, Qualtrics, Qualtrics and All Other Qualtrics Product or Service Names are Registered Trademarks or Trademarks of Qualtrics, Provo, UT, USA, 2016, ⟨http://www.qualtrics.com⟩, (n.d.).

[36] D. Fleisig, Adding information may increase overconfidence in accuracy of knowledge retrieval, Psychol. Rep. 108 (2011) 379–392, https://doi.org/10.2466/04.11.23.PR0.108.2.379-392

[37] S. Plous, The Psychology of Judgement and Decision Making, McGraw-Hill, Inc, New York, 1993.

[38] C. Janiszewski, R.S. Wyer, Content and process priming: a review, J. Consum. Psychol. 24 (2014) 96–118, https://doi.org/10.1016/j.jcps.2013.05.006

[39] K.A. Martire, R.I. Kemp, Considerations when designing human performance tests in the forensic sciences, Aust. J. Forensic Sci. 50 (2016) 166–182, https://doi.org/10.1080/00450618.2016.1229815

[40] S.C. Evans, M.C. Roberts, J.W. Keeley, J.B. Blossom, C.M. Amaro, A.M. Garcia, C.O. Stough, K.S. Canter, R. Robles, G.M. Reed, Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies, Int. J. Clin. Heal. Psychol. 15 (2015) 160–170, https://doi.org/10.1016/j.ijchp.2014.12.001

[41] M. Page, R. Lain, R. Kemp, J. Taylor, Validation studies in forensic odontology – part 1: accuracy of radiographic matching, Sci. Justice 58 (2018) 185–190, https://doi.org/10.1016/j.scijus.2017.11.001

[42] I.E. Dror, Letter to editor: on proper research and understanding of the interplay between bias and decision outcomes, Forensic Sci. Int. 191 (2009) 17–18, https://doi.org/10.1016/j.forsciint.2009.03.012

[43] I.E. Dror, Biases in forensic experts, Science 360 (80) (2018) 243, https://doi.org/10.1126/science.aat8443

[44] I.E. Dror, M.L. Pierce, ISO standards addressing issues of bias and impartiality in forensic work, J. Forensic Sci. 65 (2020) 800–808, https://doi.org/10.1111/1556-4029.14265

[45] S.M. Kassin, I.E. Dror, J. Kukucka, The forensic confirmation bias: problems, perspectives, and proposed solutions, J. Appl. Res. Mem. Cogn. 2 (2013) 42–52, https://doi.org/10.1016/j.jarmac.2013.01.001

[46] A. Camilleri, D. Abarno, C. Bird, A. Coxon, N. Mitchell, K. Redman, N. Sly, S. Wills, E. Silenieks, E. Simpson, H. Lindsay, A risk-based approach to cognitive bias in forensic science, Sci. Justice 59 (2019) 533–543, https://doi.org/10.1016/j.scijus.2019.04.003

[47] S. Nakhaeizadeh, R.M. Morgan, C. Rando, I.E. Dror, Cascading bias of initial exposure to information at the crime scene to the subsequent evaluation of skeletal remains, J. Forensic Sci. 63 (2018) 403–411, https://doi.org/10.1111/1556-4029.13569

[48] J. Klayman, Varieties of confirmation bias, Psychol. Learn. Motiv. 32 (1995) 385–418.

[49] I.A. Pretty, R.J. Pretty, B.R. Rothwell, D. Sweet, The reliability of digitized radiographs for dental identification: a web-based study, J. Forensic Sci. 48 (2003) 1325–1330.

[50] I. Dawidson, Case reports and background: difficulties with identification – Sweden, J. Forensic Odontostomatol. 29 (2011) 44–50.

[51] J. Kim, Scenarios in information seeking and information retrieval research: a methodological application and discussion, Libr. Inf. Sci. Res. 34 (2012) 300–307, https://doi.org/10.1016/j.lisr.2012.04.002

[52] M.J. Saks, Concerning L.J. Hall, E. Player, "Will the introduction of an emotional context affect fingerprint analysis and decision-making?" [Forensic Sci. Int. 181 (2008) 36-39], Forensic Sci. Int. 191 (2009) 19, https://doi.org/10.1016/j.forsciint.2009.06.011

[53] I.W. Evett, C.E.H. Berger, J.S. Buckleton, C. Champod, G. Jackson, Finding the way forward for forensic science in the US—a commentary on the PCAST report, Forensic Sci. Int. 278 (2017) 16–23, https://doi.org/10.1016/j.forsciint.2017.06.018

[54] G.S. Morrison, D.H. Kaye, D.J. Balding, D. Taylor, P. Dawid, C.G.G. Aitken, S. Gittelson, G. Zadora, B. Robertson, S. Willis, S. Pope, M. Neil, K.A. Martire, A. Hepler, R.D. Gill, A. Jamieson, J. de Zoete, R.B. Ostrum, A. Caliebe, A comment on the PCAST report: skip the "match"/"non-match" stage, Forensic Sci. Int. 272 (2017) e7–e9, https://doi.org/10.1016/j.forsciint.2016.10.018

[55] K.N. Ballantyne, J. Bunford, B. Found, D. Neville, D. Tayor, G. Wevers, D. Catoggio, An introductory guide to evaluative reporting, 2017.

[56] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A hierarchy of propositions: deciding which level to address in casework, Sci. Justice 38 (1998) 231–239, https://doi.org/10.1016/S1355-0306(98)72117-3

Chapter 6

Publication

Chiam S L, Louise J, Higgins D. "Identified", "probable", "possible" or "exclude": The influence of task irrelevant information on forensic odontology identification opinion. Science & Justice 2022;62:461–70. https://doi.org/10.1016/j.scijus.2022.06.002.

Statement of Authorship

Title of Paper	"Identified", "Probable", "Possible" and "Exclude", deciding the identification in forensic odontology. Does task-irrelevant context information matter?
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Chiam S-L, Louise J, Higgins D. "Identified", "probable", "possible" or "exclude": The influence of task-irrelevant information on forensic odontology identification opinion. Sci Justice [Internet]. 2022 Jul;62(4):461–70. Available from: https://doi.org/10.1016/j.scijus.2022.06.002

Principal Author

Name of Principal Author (Candidate)	Sher-Lin Chiam		
Contribution to the Paper	Conception and design of the project. Acquiring of research data. Analysis and interpretation of research. Wrote manuscript Acted as the corresponding author		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	20th Aug 2022

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Jennie Louise		
Contribution to the Paper	Analysis and interpretation of research		
Signature		Date	21st Aug 2022

Name of Co-Author	Denice Higgins		
Contribution to the Paper	Supervised development of work, Helped in data interpretation Manuscript evaluation and editing		
Signature		Date	18th September 2022

Science & Justice 62 (2022) 461–470

Contents lists available at ScienceDirect

Science & Justice

journal homepage: www.elsevier.com/locate/scijus

Research Paper

"Identified", "probable", "possible" or "exclude": The influence of task-irrelevant information on forensic odontology identification opinion

Sher-Lin Chiam [a,*], Jennie Louise [b], Denice Higgins [a]

[a] Forensic Odontology Unit/ Adelaide Dental School. The University of Adelaide, Australia
[b] Faculty of Health Sciences. The University of Adelaide, Australia

ARTICLE INFO

Keywords:
Forensic odontology
Human identification
Context bias
Contextual information
Forensic odontology identification categories

ABSTRACT

In a mass disaster situation, identification of the deceased utilising comparison of dental features is frequently heavily relied upon to facilitate rapid and accurate outcomes. The method consists of the comparison of clinical and radiographic records depicting oral structures and dentition to allow an opinion to be produced on a presumed identity. Current forensic odontology identification opinions are expressed as categories of levels of identification. Categories such as "Identified", "Probable", "Possible" and "Exclude" are used in various forensic odontology identification scales. The boundaries between the levels of the scales are not fixed; hence, category selection is highly subjective. It is uncertain how extrinsic factors such as exposure to contextual task-irrelevant information or operator experience influence category selection. In this study, forensic odontologist and dentist participants read task-irrelevant context case information containing either strong or weak identification or non-identification suggestions before evaluating and comparing pairs of true matching and non-matching dental radiographs. They were then asked to form an opinion regarding identification using one of four categories from the INTERPOL scale. Context information was found to influence categorical decisions. The magnitude and direction of influence depended on the type of participant, the true match status of the radiographs, and the strength and direction of bias of the context. The results of this study demonstrate the contextual effect and fluidity of the boundaries between the categories on the identification scale and highlight the need for stringent protocols to be developed regarding the use of these categorical scales to enable decision making to be more objective.

Task irrelevant context information has been shown to affect expert decision making in multiple disciplines of forensic science [1–6] including forensic odontology [7]. In this previous study [7], participants were exposed to task-irrelevant information, then compared pairs of dental radiographs and rated the probability of the radiographs being from the same person on a continuous rating scale. Strong task-irrelevant context information was shown to affect the accuracy and probabilistic matching evaluation in the direction of context suggestion. In practice, however, opinion is explicitly expressed using a scale with categories of certainty of the identification not as a match probability on a continuous scale. Therefore, whether irrelevant context affects the final forensic odontology opinion on identification in real cases remains unknown.

The likelihood or proposition approach is not routinely used for forensic odontology identification opinions. Most practitioners adopt categorised scales to state their opinions. Various recognised standardised scales are used [8] including the American Board of Forensic Odontology (ABFO) [9], and the International Criminal Police Organisation (INTERPOL) disaster victim forensic odontology identification scale [10,11]. The number of categories and terminology used may differ slightly between scales, but all scales are framed by terms connoting definitive identification or exclusion interspersed with categories expressing varying levels of lower certainty, for example: probable or possible. These categories are believed to represent confidence in the identification, which has been partially verified by the correlation of self-reported confidence level with the semantic certainty of the categories demonstrated in a previous study [12]. However, it is uncertain whether the final category choice is solely driven by the probability judgment of the match or if additional considerations are included because the evidential weight evaluation process is not elucidated and there are no explicit rules to guide the rationale of category choice [13,14]. Therefore, it remains unclear what factors inform the selection

* Corresponding author.
E-mail address: sher-lin.chiam@adelaide.edu.au (S.-L. Chiam).

https://doi.org/10.1016/j.scijus.2022.06.002
Received 13 May 2022; Received in revised form 24 June 2022; Accepted 30 June 2022
Available online 2 July 2022

of category or what information is considered consciously and/or unconsciously.

If category selection is based solely on match probabilistic judgment, then context will impact category decision as probability judgment has been demonstrated to be influenced by context [7]. However, categorisation requires imposing boundaries on a continuum judgement [15,16]. These threshold boundaries are susceptible to many external and internal contextual and affective factors thus, category membership decisions are malleable and unstable [17]. Context effects may arise from case information, human factors in the work environment [18], and even microcosmic contextual nuances of rating scales. Rating scales frame the decision choices and therefore require a threshold range to be assigned to each category on the scale, and these may not be evenly distributed [19,20]. Factors including the scale design or granularity, connotation, hierarchy of categories [19,21–23], and practitioner interpretation, can affect this threshold distribution. It has been demonstrated that the position and number of available categories on the scale affect the fluidity of interpretation and application [12]. In this previous work [12], it was noted that the term "Possible" was seldom associated with non-identification when an additional category "Insufficient information" was available and positioned between the terms "Possible" and "Exclude". However, when the term "Possible" is the immediate term before "Exclude", its selection represents both identification and non-identification beliefs. Hence, the intermediate terms may have wide threshold boundaries, implying that the choice of category may remain unchanged despite changes in the internal decision process. The general cognitive task of categorisation increases the forensic odontology identification decision load and, thereby, adds to the complexity of the effect of contextual bias on the decision process. The complex effects of contextual bias on decisions can be further compounded by potential variation in the effect of contextual information and interpretation of the scale between expert and non-expert practitioners. Although context information was shown to affect the match probabilistic belief in the reference study, [7] it is uncertain whether this also applies to category selection.

Paradoxically, experts have been theorised to be more prone to context effects because of their use of implicitly and explicitly learned cognitive strategies. Experts often chunk a large amount of information into meaningful summarised groupings and apply a top-down process, which involves working from predicted outcomes based on similar past cases encountered [24,25]. This cognitive advantage in expertise may also potentially induce unconscious incorporation of contextual information changing the evidential weight evaluation process. Evaluation of the weight of evidence from radiographs requires a firm foundation in dental knowledge. This knowledge allows recognition of iatrogenic and biological features and allows the practitioner to account for differences in radiographs of the same person due to radiographic artefacts. This ability is acquired through implicit learning and experience through the practice of dentistry and therefore is a common set of skills between dentists and forensic odontologists [26,27]. An integral part of the expertise in forensic odontology is understanding the legal implications of the levels of identification, which surpasses the nuances of the lay semantic understanding of the terms on the scales. Some of these implications include the cost of different types of errors, for example, false identification versus false exclusion. Formal training or apprenticeship would familiarise the forensic odontologist with an appreciation and understanding of the legal implications of the levels [28,29]. Although it has been shown that there is no significant difference in the effect of context on the judgment of match or non-match for forensic odontology and dentist participants [7]familiarity with the scales and the legal implications of the final decision may result in a difference between the groups when assigning categories on the scale.

This study explores the effect of context information on the final opinion of identification expressed as a category on a forensic dental scale formed by forensic odontologists and dentists. The scale chosen in this study is recommended by INTERPOL for forensic odontology identification in disaster victim identification, with the intent to allow communication and understanding from an international perspective.

1. Materials and methods

1.1. Experimental design

The aspects pertinent to this study are highlighted below, for more details on this inter-subject experiment, see [7].

1.2. Materials

The participants were 26 practicing forensic odontologists and 22 dentists (N = 48) from various countries (Australia, Finland, India, Italy, Norway, South Africa, Sweden, New Zealand, and the Netherlands). 12 dentists had more than 10 years of general dentistry practice experience (range: less than 5 to more than 20 years). 3 had a graduate diploma or higher in forensic odontology, and 6 attended non-award short courses as part of continuing professional development; however, none of the dentists practiced forensic odontology. 14 of the forensic odontologists had more than 10 years of experience (range: less than 5 to more than 20 years). Regarding forensic odontology education, 4 had attended non-award courses and 22 had a graduate diploma or higher. 14 were registered as specialists in forensic odontology.

1.3. Stimuli (N = 12)

- 6 pairs of matching radiograph pairs (Confirmed by forensic odontology and DNA identification.)
- 6 pairs of non-matching radiographs
- Each pair was appended with two types of context information of the same strength but of opposing identification cues to the actual ground truth (i.e., either strong supportive and strong contradictory, or weak supportive and weak contradictory). The difference in magnitude between the supportive and contradictory effect between the strong and weak context groups quantified the effect of the strength of the context.
- Therefore, 3 of the matching and 3 of the non-matching radiograph pairs were given strong contexts while the remainder were given weak contexts to ensure a balance of the test cases.

1.4. Contextual information

The contextual information consisted of hypothetical scenarios devoid of obvious task-relevant information (e.g., treatment history, dates of radiograph exposure) to simulate the possible bias induced by contact with police circumstantial case reports.

The manipulation in the vignettes was the strength and direction of bias evoked through the application of theories regarding the human cognitive architecture when interacting, assimilating, and processing information [30–35]. Contextual identification and exclusion suggestions framed and anchored the direction of bias.

Strong suggestions were induced by the consistency, coherence, and quantity of definitive information in the narrative [36–38], and clear first and last impression (the primacy and recency effect) [34,39] by placing definitive information at the beginning and end of the narratives. Commonly encountered cases were used as models for the vignettes to increase credibility (an important factor in the use of scenarios in a test environment).

Weak suggestions were designed with the introduction of doubt and uncertainty in the direction of bias and therefore lack a very clear and definitive sense of direction of bias.

Supplementary elements that controlled the strength included suggestions of the cost of incorrect decisions versus routine unsuspicious cases, for example, murder cases.

Non-test vignettes with radiograph exposure dates provided were

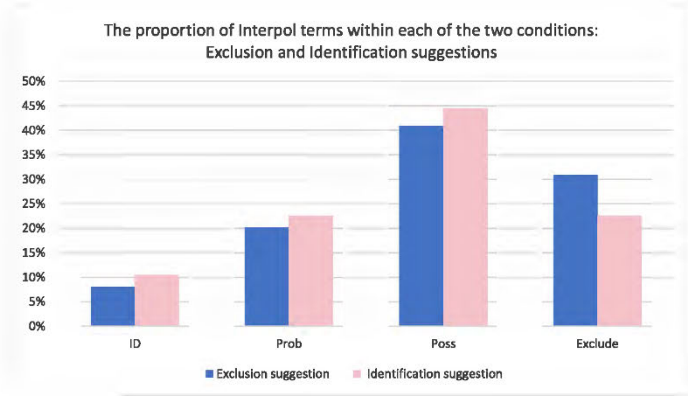


Fig. 1. The proportion of the INTERPOL terms within each context group: Exclusion (n = 288 decisions) and Identification (n = 288 decisions) suggestions.

included to prevent the participants from learning that the test vignettes did not have relevant information and to encourage reading through all the vignettes.

Contact with the information was also encouraged by requiring participants to actively reveal and progress through the paragraphs to completion before evaluating the radiographs.

The vignettes induced possible resultant cognitive bias, by altering, modulating, and anchoring the prior expectation, the tendency for confirmation, and overall confidence associated with judgment and decision before the evaluation and comparison of the radiograph pairs.

There were four possible types of scenarios: strong match, strong non-match, weak match, and weak non-match, for pairing with the 12 dental radiograph sets.

A total of 24 vignettes were presented, two for each pair of radiographs; each pair of radiographs was provided with contrasting direction of bias of the same strength, which means that each participant viewed 12 cases and vignettes.

1.5. Method

The experiment was delivered online using Qualtrics ™ software [40]. Participants were randomly assigned to either of two groups. All participants had to watch an instructional video and attempt a practice case before proceeding to the actual trial. Both groups viewed the same set of radiographs, but the context information differed in the direction of bias only but not the strength of suggestion. The order of the cases was randomised for each participant.

For each case, the participants had to read circumstantial information before viewing and comparing a pair of dental radiographs. The time spent reading the information was automatically recorded.

Then they had to indicate their opinion of the identification by choosing one of the four possible terms on the INTERPOL scale (Version 2009) [11], the definition of each term was accessible by hovering the pointer over the terms before choice. After choosing a category, participants were asked if they would have chosen the category "Insufficient evidence" if it was available (Yes or No answer).

INTERPOL terms and definitions [11]:

- Identified (ID):

There is absolute certainty the PM and AM records are from the same person.

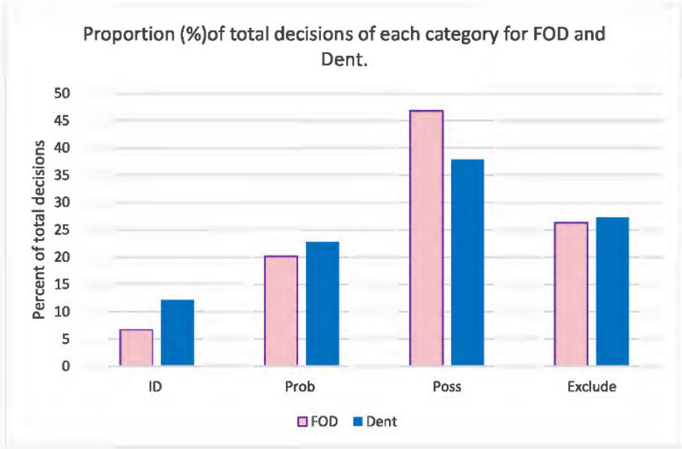
- Probable (Prob):

Specific characteristics correspond between PM and AM but either PM or AM data or both are minimal.

- Possible (Poss):

Table 1
The count and proportion of the INTERPOL terms within each context group for match and non-match radiograph pairs.

	Categories	Strong Supportive (% within column)	Strong Contradictory (% within column)	Weak Supportive (% within column)	Weak Contradictory (% within column)	Total (% within column)
Match X-rays	ID	16 (21.6)	10 (14.3)	8 (11.4)	11 (14.9)	45 (15.6)
	Prob	24 (32.4)	17 (24.3)	24 (34.3)	30 (40.5)	95 (33.0)
	Poss	21 (28.4)	27 (38.6)	30 (42.9)	25 (33.8)	103 (35.8)
	Exclude	13 (17.6)	16 (22.9)	8 (11.4)	8 (10.8)	45 (15.6)
	Total	74	70	70	74	288
Non-match X-rays	ID	1 (1.4)	4 (5.7)	1 (1.4)	2 (2.7)	8 (2.8)
	Prob	7 (9.5)	10 (14.3)	4 (5.7)	7 (9.5)	28 (9.7)
	Poss	27 (36.5)	36 (51.4)	39 (55.7)	41 (55.4)	143 (49.7)
	Exclude	39 (52.7)	20 (28.6)	26 (37.1)	24 (32.4)	109 (37.8)
	Total	74	70	70	74	288
	Total count of decisions	148	140	140	148	576



Identification	FOD N (%)	Dent N (%)
ID	21 (6.73)	32 (12.12)
Prob	63 (20.19)	60 (22.73)
Poss	146 (46.79)	100 (37.88)
Exclude	82 (26.28)	72 (27.27)

Fig. 2. The proportion of each INTERPOL term within the total number of decisions in the Forensic Odontologist (FOD) and Dentist (Dent) participant groups.

There is nothing that excludes the identity but either PM or AM data or both are minimal.

- Exclude (Exclude):

PM and AM records are from different persons.

- No comparison can be made

The category "No comparison can be made" was NOT included as a choice on the scale.

2. Results

Decision direction appears to align with the direction of bias. This effect is greater in the strong contextual information group than in the weak group. Fig. 1 displays the frequency of each INTERPOL term used under the influence of match and non-match contextual suggestions. This reveals that ID, Prob and Poss choice rates were higher with suggestion of match while Exclude was higher with suggestion of non-match. Decision alignment with the direction of strong bias was observed when the data were stratified by radiograph match status, strength, and direction of contextual information (Table 1).

A general survey into the overall frequency of choice of category showed a difference between forensic odontology and dentist participants. (Fig. 2). ID was used half as often while Poss was used 1.3 times more frequently in FOD compared to Dent.

To explore if participants and radiograph pair types exerted an interactive effect, the frequencies for each INTERPOL term were charted for each participant group in the four different context conditions stratified by the true match status of the radiograph pairs (Figs. 3a–d).

For true non-matching radiographs, the rate of incorrect identification for the FOD was 3% (1 out of 36 decisions) and only occurred with strong contradictory (match) context. For Dent, the incorrect identification rate was 8% (3 out of 34 decisions) under strong contradictory (match) context and 3% under strong supportive (non-match) context (Fig. 3a).

Exclude (correct exclusions) rates increased by 1.8 and 8 times with strong supportive (non-match) suggestions compared to contradictory (match) suggestions for the FOD and Dent respectively (Fig. 3a). Correct exclusion rates were 1.5 times higher with weak supportive (non-match) context for the Dent but comparable for the FOD (Fig. 3b).

For true matching radiographs, the frequency of correct identification increased by approx. 9 times (Fig. 3c) with strong supportive (match) suggestions for FOD only, while wrong exclusion rate was higher (approx. 1.6 times) for Dent with strong contradictory (non-match) context compared to supportive (Fig. 3c). Interestingly, for the Dent, the correct identification rate was 3 times higher with weak contradictory (non-match) than with supporting context (match) (Fig. 3d).

As the summarised data (Figs. 1-3, and Table 1) suggests that the effect of contrasting context on the decision outcomes depended on the participants and radiograph pair types, a mixed-effects ordinal logistic regression model was fitted, with random effects for participant and radiograph pair. A three-way interaction between radiograph type (true identified vs exclude), professional type (FOD vs Dentist) and supportive vs contradictory information was included. Odds ratios of being in a higher-ordered category: 1 = ID, 2 = Prob, 3 = Poss and 4 = Exclude with supportive evidence vs contradictory evidence were calculated. An estimate less than 1 = lower odds of choosing a higher category (e.g., Exclude vs Poss/Prob/ID) with supportive vs contradictory evidence, and an estimate greater than 1 = higher odds.

These data suggest that the effect of strong supportive versus contradictory information differed between the FOD and the Dent groups (P

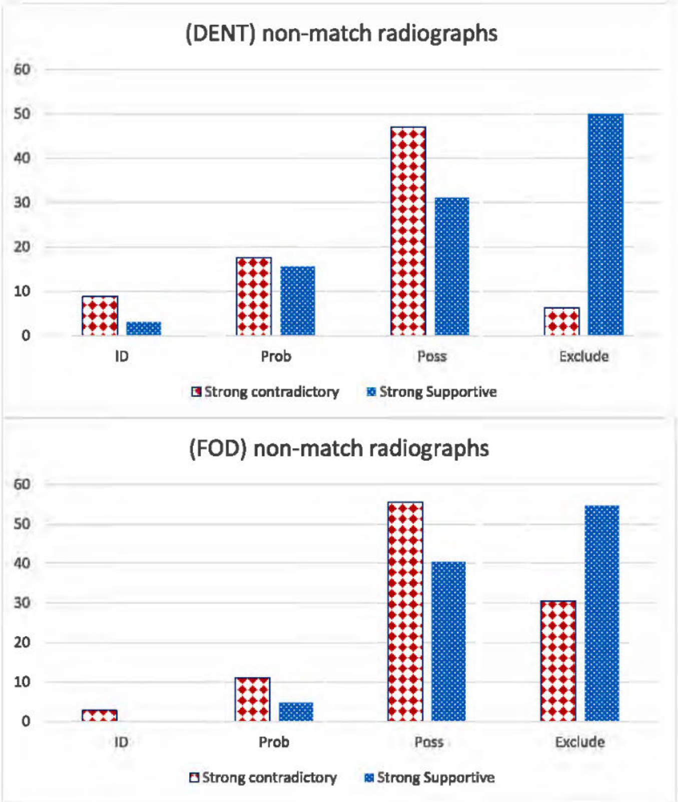


Fig. 3a. Comparison of the frequency (%) of each term within each strong contradictory and supportive context group for non-matching radiographs for dentist (Dent) and forensic odontologist (FOD) participants.

= 0.099) for true match radiograph pairs. Although this effect was not statistically significant, the odds of choosing a higher category (towards exclusion) were lower for FOD (OR 0.62, 95% CI 0.34, 1.14) but not for Dent (OR 1.33; 95% CI 0.55, 3.24). This suggests that strong supportive information encouraged choosing categories towards higher certainty of identification for the FOD while there was no real evidence of an effect for Dent. For true non-match radiograph pairs, supportive information increased the odds of choosing a higher category (towards exclusion) for both FOD and Dent, but the effect was stronger in Dent. For FOD the OR was 1.51 (95% CI 0.82, 2.77, $p = 0.190$) while for Dent it was 3.00 (95% CI 1.49, 6.03, $p = 0.002$).

To examine the effect of strength of suggestion on the odds of choosing the category of ID above all others for true matching cases a mixed model was employed where the outcomes were dichotomised as the category ID versus all other categories. Radiograph pair type was included as a fixed rather than a random effect. Separate models were fitted for strong supportive vs contradictory, and weak supportive vs contradictory. Strong supportive information significantly increased the odds of choosing ID by approximately 14 times (95% CI 1.23, 153.48, $p = 0.033$) when compared to strong contradictory information for the FOD, although statistically significant it was also noted that the small sample size resulted in an extremely wide confidence interval. For the dentist participants, this effect was not seen and the difference between the two participant groups was significant at $p = 0.038$.

Weak supportive information had a similar but non-significant effect for the FOD but for the DEN, it had a marginally significant effect of decreasing the odds of choosing ID by 0.08 times (95% CI 0.00, 1.35, $p = 0.079$). However, once again these results are limited by the wide confidence interval associated with the small data size.

Table 2 shows the results for "Would you have chosen the category 'Insufficient evidence' if it was available". For those FOD who chose ID, none indicated that they would have chosen "Insufficient", conversely, 18.75% of the Dent group indicated they would have. Poss was the category with the highest frequency of being substituted with "Insufficient Information" if it had been an available option at approximately 79% and 82% for the FOD and Dent respectively. For all categories except for when they had chosen Exclude, the Dent group would have chosen "Insufficient" more often than the FOD group.

More than 70% of all the participants had above 30 sec of contact time with the each of the vignettes.

3. Discussion

This study investigated the effect of context information on the selection of opinion category on the INTERPOL scale for general dentists and forensic odontologists when comparing dental radiographs to

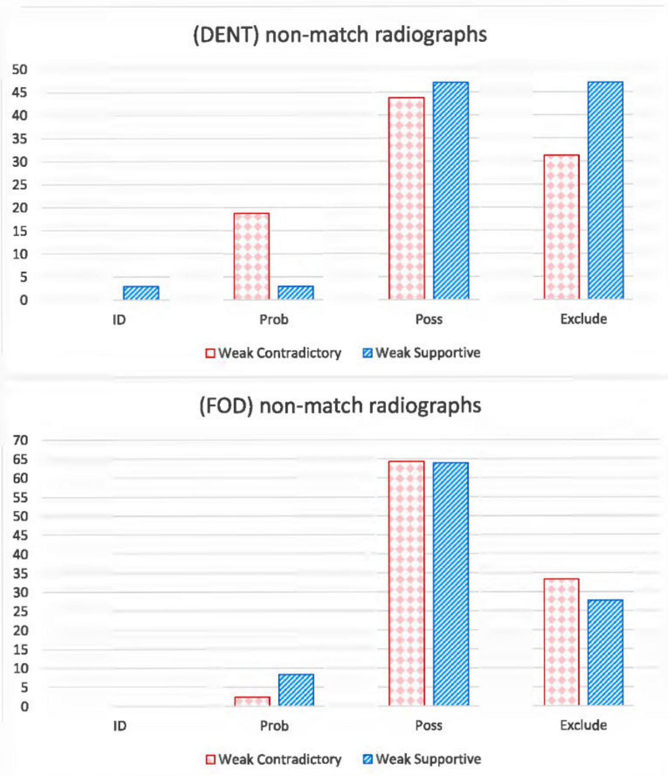


Fig. 3b. Comparison of the frequency (%) of each term within each weak contradictory and supportive context group for non-matching radiographs for dentist (Dent) and forensic odontologist (FOD) participants.

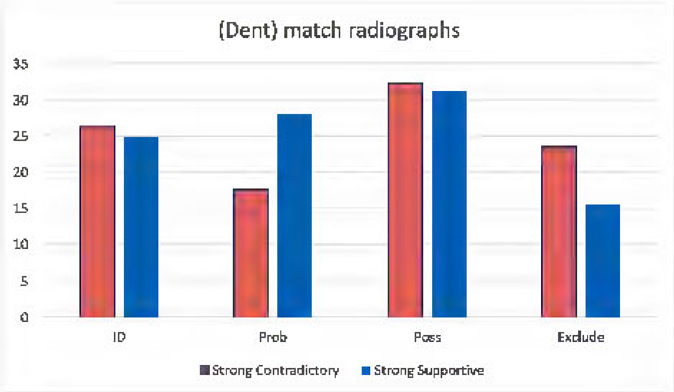
determine identification.

It appeared that the forensic odontologist group was more conservative in their decisions than dentists when confirming identity as indicated by the lower frequency of the use of ID and higher use of Poss. However, this was not true for excluding identity with comparable use of Exclude between groups. This indicates a possible systematic difference between the two groups. The difference between the participant groups is unlikely due to contact time with the vignettes, as no significant differences were noted regarding contact time.

Overall, there is evidence of alignment of decision direction with the direction of bias for the forensic odontologist participants. With the effect being more pronounced and consistent with strong context suggestions. Strong supportive context was associated with higher correct identification and exclusion rates for the forensic odontologist participants. For dentist participants, strong suggestion was associated with increased correct exclusion but was also associated with increased wrong identification. Interestingly, weak contradictory context seemed to increase correct identification in this group.

Context effect is unequivocally demonstrated when the decision changes align with the suggested bias direction. In instances where decision changes are opposite to the suggested bias direction, the biasing effect becomes debatable; this was found predominately in dentist participants for weak context. Possible reasons for such results could be the small sample size or misinterpretation of the context suggestion by these participants. The weak context required juxtaposing varying amounts of identification and exclusion suggestions to modulate the strength which could have confused the intended direction of the suggestion. Interestingly, this effect was less observed amongst the forensic odontologists, this difference may rest on efficiency in assimilating the information in the vignettes. As the context information was modelled on actual commonly encountered scenarios; experience, practice, and familiarity may have increased the speed and ability to grasp the overall essence of the suggestions for the forensic odontologist participants. This enhanced efficiency at managing information although a hallmark of expertise, could have also led to unconsciously factoring the information into their decisions resulting in a more pronounced relationship of context and decisions [25,31,34].

Paradoxically experts have been posited to possibly be more susceptible to context effect [24,25]. Experience and training lead to a more heuristic or top-down approach to decision making which increases the efficiency in integrating information for evaluation of the evidence [25,35,41]. This may also predispose the unconscious incorporation of contextual information into the evaluation of evidence. It is perhaps this seemingly improved decision making that has contributed to intransigent arguments and beliefs about the role of context information even after a decade of research, for example by authors Curley et al. 2000 [42]. One possible reason may be the confusion about the definition of contextual information as delineated by Thompson 2020 [43] in his



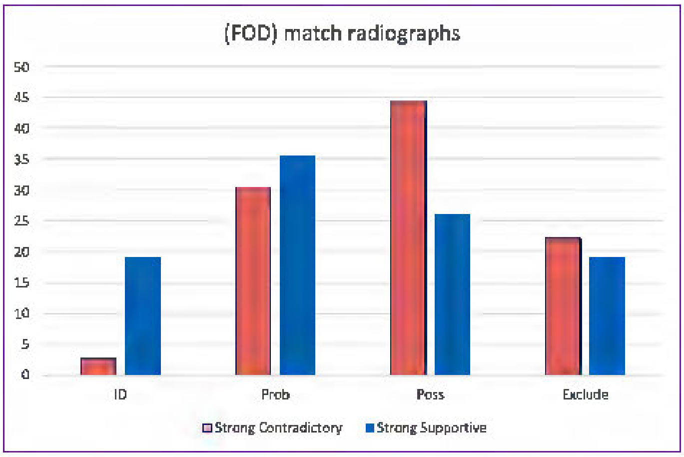


Fig. 3c. Comparison of the frequency (%) of each term within each strong contradictory and supportive context group for matching radiographs for dentist (Dent) and forensic odontologist (FOD) participants.

comment, citing the clear position and recommendation by the National Commission on Forensic Science [44] that contextual information used in forensic disciplines refers to task-irrelevant information. A further related consideration is not only the concept of context information but also the need for consensus regarding what constitutes relevant and irrelevant information. Task irrelevant information may increase the certainty of decisions, but using such information violates the principle of independence of evidence consideration which undermines the value of expert opinion and is inconsistent with true impartiality and justice [43,45,46].

More importantly, the bias in contextual information may potentially lead to erroneous decisions. Determining the accuracy of the decision is only possible for the two definitive terms in the scale, ID and Exclude. The intervening terms allow room for the expression of uncertainty about the judgment. Wrong decisions or errors in this experiment were accessible due to the known ground truth of the radiograph pairs. Stratified analysis by the radiograph pair type showed an interactive effect indicating that the context effect was different for the two decision types: match and non-match. While the context was not associated with a significantly increased rate of incorrect identification for the forensic odontology group, this effect was observed in the dentists. This is an interesting finding especially given that it was shown that accuracy for both groups of participants was equally affected by strong context when forced to make binary match and non-match decisions for the same radiograph pairs [7]. It is hence possible that a difference lies in the understanding of the legal implications for the categories particularly ID. It should be noted that the definition for each term in the scale in the experiment was available before choosing in every case, therefore it does appear to be more the interpretation of the implication, not just the definition. Further evidence of a difference in awareness of the legal significance of categories can be seen in the overall lower frequency of use of ID by the forensic odontology group and unanimous rejection of the alternative category "insufficient information" when ID had been chosen by this group.

Encompassed in the forensic evidentiary implication of ID is also the understanding of the cost of different types of error for the forensic odontology group. Incorrect exclusion although a definitive decision is comparatively viewed as less serious, than false identification [27,47]. This may explain why the false exclusion rate was comparable between the two groups.

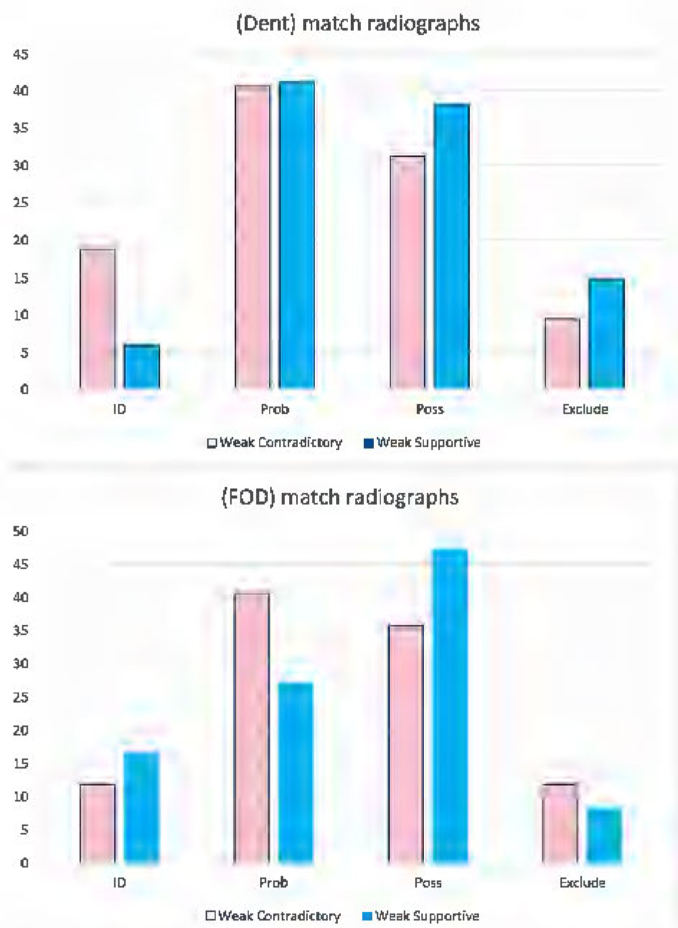


Fig. 3d. Comparison of the frequency (%) of each term within each weak contradictory and supportive context group for matching radiographs for dentist (Dent) and forensic odontologist (FOD) participants.

Table 2
Percentage and number of instances where the participants would have chosen the category 'Insufficient' if it was available.

	DEN		FOD	
	Count (% in row)	Count (% in row)	Count (% in row)	Count (% in row)
	Would have chosen insufficient	Not have chosen Insufficient	Would have chosen insufficient	Not have chosen Insufficient
ID	6 (18.75)	26 (81.25)	0 (0.00)	21 (100.00)
Probable	24 (40.00)	36 (60.00)	17 (26.98)	46 (73.02)
Possible	82 (82.00)	18 (18.00)	115 (78.77)	31 (21.23)
Exclusion	43 (59.72)	29 (40.28)	48 (58.54)	34 (41.46)
Grand Total	155 (58.71)	109 (41.29)	180 (57.69)	132 (42.31)

While strong supportive context was associated with a significantly increased rate of correct exclusion for both groups of participants, there was also increased incorrect exclusion with strong contradictory context. Exclusion is a rare event for forensic odontologists in routine coronial casework and the lack of practice would likely mean less confidence in making such a decision [28]. Context information might have influenced borderline difficult decisions which is consistent with how the context effect is often seen more prominently in difficult decisions. Another reason for the more pronounced effect in exclusion decisions may relate to the semantic effect of the terms in the scale itself. Apart from Exclude, the other mid-level terms appear to express the probability that the records or radiographs are of the same person; especially as in actual practice, most comparisons are of believed to be persons, rarely is an exclusion specifically requested. Therefore, where there is an inclination towards exclusion, the term Poss may not be perceived as the most appropriate expression of a sub-level of exclusion, this preference and reservation to use the term Poss was also noted in another study [12]. Under this condition, a context that suggests exclusion would increase the confidence and support commitment to the category Exclude.

The control on variables and the research design with consideration for human performance [48] were measures taken to ensure internal validity, however, the small sample size and lab-based test conditions restrict and limit the generalisability and ecological applicability of our observations.

This exploratory study demonstrated that irrelevant context information impacted the final opinion, it changed the certainty and increased the use of more definitive levels of identification in line with the suggestions. It thus demonstrated the context dependant fluidity in the application of the categories, but importantly there was no evidence of increased incorrect identification for the forensic odontologist participants.

CRediT authorship contribution statement

Sher-Lin Chiam: Conceptualization, Methodology, Investigation, Writing – original draft, Data curation, Formal analysis. **Jennie Louise:** Formal analysis, Writing – review & editing. **Denice Higgins:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] I.E. Dror, G. Hampikian, Subjectivity and bias in forensic DNA mixture interpretation, Sci. Justice 51 (2011) 204–208, https://doi.org/10.1016/j.scijus.2011.08.004.

[2] I.E. Dror, A.E. Péron, S.L. Hind, D. Charlton, When emotions get the better of us: the effect of contextual top-down processing on matching fingerprints, Appl. Cognit. Psychol. 19 (2005) 799–809, https://doi.org/10.1002/acp.1130.

[3] N.K.P. Osborne, S. Woods, J. Kieser, R. Zajac, Does contextual information bias bitemark comparisons? Sci. Just. 54 (2014) 267–273, https://doi.org/10.1016/j.scijus.2013.12.005.

[4] I. Dror, J. Melinek, J.L. Arden, J. Kukucka, S. Hawkins, J. Carter, D.S. Atherton, Cognitive bias in forensic pathology decisions, J. Forensic Sci. 66 (2021) 1751–1757, https://doi.org/10.1111/1556-4029.14697.

[5] H.J. Hamnett, I.E. Dror, The effect of contextual information on decision-making in forensic toxicology, Foren. Sci. Int. Synergy 2 (2020) 339–348, https://doi.org/10.1016/j.fsisyn.2020.06.003.

[6] S. Hartley, A.P. Winburn, I.E. Dror, Metric forensic anthropology decisions: reliability and biasability of sectioning-point-based sex estimates, J. Forensic Sci. 67 (2022), https://doi.org/10.1111/1556-4029.14931.

[7] S.L. Chiam, I.E. Dror, C.D. Huber, D. Higgins, The biasing impact of irrelevant contextual information on forensic odontology radiograph matching decisions, Forensic Sci. Int. 327 (2021), 110997, https://doi.org/10.1016/j.forsciint.2021.110997.

[8] D. Higgins, H. James, Classifications used by Australian forensic odontologists in identification reports, J. Foren. Odonto-Stomatol. 24 (2006) 32–35.

[9] American Board of Forensic Odontology, ABFO Body Identification Guidelines, (n.d.). http://abfo.org/wp-content/uploads/2012/08/ABFO-Body-ID-Information-Guidelines-Feb-2017.pdf (accessed June 18, 2017).

[10] INTERPOL Disaster Victim Identification Guide (Version 2014), (2014). https://www.interpol.int/content/download/589/file/18Y1344 E DVI_Guide.pdf (accessed June 19, 2017).

[11] INTERPOL Disaster Victim Identification Guide (Version 2009), (2009) 1–55. https://www.cmu.edu/chrs/conferences/eppi/docs/Interpol DVI Guide.pdf (accessed June 15, 2019).

[12] S.L. Chiam, D. Higgins, K. Colyvas, M. Page, J. Taylor, Interpretation, confidence and application of the standardised terms: identified, probable, possible, exclude and insufficient in forensic odontology identification, Sci. Justice 61 (2021) 426–434, https://doi.org/10.1016/j.scijus.2021.02.007.

[13] L. Gorza, S. Mânica, Accuracy of dental identification of individuals with unrestored permanent teeth by visual comparison with radiographs of mixed dentition, Foren. Sci. Int. 289 (2018) 337–343, https://doi.org/10.1016/j.forsciint.2018.06.004.

[14] S. Martin-de-Las-Heras, A. Valenzuela, J.D.D. Luna, M. Bravo, The utility of dental patterns in forensic dentistry, Foren. Sci Int. 195 (166) (2010) e1–e5, https://doi.org/10.1016/j.forsciint.2009.11.004.

[15] S. Harnad, Categorical Perception, in: Encyclopedia of Cognitive Science, Nature Publishing Group: Macmillan, 2003. 10.1002/0470018860.s00490.

[16] I.E. Dror, G. Langenburg, " Cannot Decide " : the fine line between appropriate inconclusive determinations versus unjustifiably deciding not to decide *, J. Forensic Sci. 64 (2019) 10–15, https://doi.org/10.1111/1556-4029.13854.

[17] M.W. Eysenick, M.T. Keane, Objects, Concepts, and Categories, in: Cognitive Psychology : A Student's Handbook, Taylor & Francis group, 2013: pp. 306–334.

[18] I.E. Dror, Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias, Anal. Chem. 92 (2020) 7998–8004, https://doi.org/10.1021/acs.analchem.0c00704.

[19] T.C. Brown, T.C. Daniel, Scaling of ratings: concepts and methods, USDA Forest Service Research Paper. RM-293 (1990) 1–24.

[20] D. Kahneman, O. Sibony, C.R. Sunstein, Noise: A flaw in human judgment, First edit, Little, Brown Spark, New York, NY, 2021.

[21] T. Amoo, H. Friedman, Do numeric values influence subjects' responses to rating scales? J. Int. Market. Market. Res. 26 (2001).

[22] H.H. Friedman, T. Amoo, Rating the rating scales, J. Market. Manage. 9 (1999) 114–123.

[23] H. Taherdoost, What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale / Attitude Scale / Likert Scale, International Journal of Academic Research in Management (IJARM). 8 (2019) 2296–1747.

[24] I. Dror, The ambition to be scientific: human expert performance and objectivity, Sci. Just. 53 (2013) 81–82, https://doi.org/10.1016/j.scijus.2013.03.002.

[25] I.E. Dror, The paradox of human expertise: why experts get it wrong, in: N. Kapur, A. Pascual-Leone, V. Ramachandran, J. Cole, S. Della Sala, T. Manly, A. Mayes, O. Sacks (Eds.), The Paradoxical Brain, Cambridge University Press, 2011, pp. 177–188.

[26] V. Pinchi, G.A. Norelli, F. Caputi, G. Fassina, F. Pradella, C. Vincenti, Dental identification by comparison of antemortem and postmortem dental radiographs: Influence of operator qualifications and cognitive bias, Foren. Sci. Int. 222 (2012) 252–255, https://doi.org/10.1016/j.forsciint.2012.06.015.

[27] A. Wenzel, A. Richards, J. Heidmann, Matching simulated antemortem and postmortem radiographs from human skulls by dental studens and experts: Testing skills for pattern recognition, J. Foren. Odontostomatol. 28 (2010) 5–12.

[28] I.A. Pretty, R.J. Pretty, B.R. Rothwell, D. Sweet, The reliability of digitized radiographs for dental identification: a web-based study, J. Forensic Sci. 48 (2003) 2002032, https://doi.org/10.1520/jfs2002032.

[29] H. Soomer, M.J. Lincoln, H. Ranta, A. Penttila, E. Leibur, A. Penttilä, E. Leibur, Dentists' qualifications affect the accuracy of radiographic identification, J. Foren. Sci. 48 (2003) 1121–1126, https://doi.org/10.1520/jfs2003142.

[30] D. Charlton, P.A.F. Fraser-Mackenzie, I.E. Dror, Emotional experiences and motivating factors associated with fingerprint analysis, J. Foren. Sci. 55 (2010) 385–393, https://doi.org/10.1111/j.1556-4029.2009.01295.x.

[31] P.A.F. Fraser-Mackenzie, R.E. Bucht, I.E. Dror, in: Comparative Decision Making, Oxford University Press, 2013, pp. 385–415.

[32] D. Kahneman, Thinking fast and slow, First Edit, Farrar, Straus and Giroux, New York, 2011.

[33] S.M. Kassin, I.E. Dror, J. Kukucka, The forensic confirmation bias: problems, perspectives, and proposed solutions, J. Appl. Res. Mem. Cognit. 2 (2013) 42–52, https://doi.org/10.1016/j.jarmac.2013.01.001.

[34] C.L. Maclean, I. Dror, A primer on the psychology of cognitive bias, in: C. T. Robertson, A.S. Kesselheim (Eds.), Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science, and Law, 1st ed., Academic Press, 2016, pp. 13–24.

[35] P.A. Zapf, I.E. Dror, Understanding and mitigating bias in forensic evaluation: lessons from forensic science, Int. J. Foren. Mental Health 16 (2017) 227–238, https://doi.org/10.1080/14999013.2017.1317302.

[36] A. Koriat, S. Lichtenstein, B. Fischhoff, B. Combs, Reasons for confidence, journal of experimental psychology: human, Learn. Mem. 6 (1980) 107–118.

[37] P.D. Kvam, T.J. Pleskac, Strength and weight: the determinants of choice and confidence, Cognition 152 (2016) 170–180, https://doi.org/10.1016/j.cognition.2016.04.008.

[38] D. Fleisig, Adding information may increase overconfidence in accuracy of knowledge retrieval, Psychol. Rep. 108 (2011) 379–392, https://doi.org/10.2466/04.11.23.PR0.108.2.379-392.

[39] S. Plous, The psychology of judgement and decision making, McGraw-Hill Inc, New York, 1993.

[40] Qualtrics software, Copyright ©, Qualtrics, Qualtrics and All Other Qualtrics Product or Service Names are Registered Trademarks or Trademarks of Qualtrics, Provo, UT, USA, 2016, http://www.qualtrics.com, (n.d.).

[41] M.W. Eysenick, M.T. Keane, Judgement and decision making, in: Cognitive Psychology : A Student's Handbook, 6th ed., Taylor & Francis group, 2013: pp. 499–532.

[42] L.J. Curley, J. Munro, M. Lages, R. MacLean, J. Murray, Assessing cognitive bias in forensic decisions: a revieW AND OUTLook, J. Forensic Sci. 65 (2020) 354–360, https://doi.org/10.1111/1556-4029.14220.

[43] W.C. Thompson, Commentary on: Curley LJ, Munro J, Lages M, MacLean R, Murray J. Assessing cognitive bias in forensic decisions: a review and outlook. J Forensic Sci doi: 10.1111/1556-4029.14220. Epub 2019 Nov 6, Journal of Forensic Sciences. 65 (2020) 666–667. 10.1111/1556-4029.14272.

[44] National Commission on Forensic Science - Human Factors Subcommittee, Ensuring That Forensic Analysis is Based Upon Task-Relevant Information, 2015.

[45] W.C. Thompson, J. Vuille, A. Biedermann, F. Taroni, The role of prior probability in forensic assessments, Front. Genet. 4 (2013) 220, https://doi.org/10.3389/fgene.2013.00220.

[46] W.C. Thompson, What role should investigative facts play in the evaluation of scientific evidence? Aust. J. Forensic Sci. 43 (2011) 123–134, https://doi.org/10.1080/00450618.2010.541499.
[47] I. Coyle, D. Field, P. Wenderoth, Pattern recognition and forensic identification: The presumption of scientific accuracy and other falsehoods, Crim. Law J. 33 (2009) 214–226.
[48] K.A. Martire, R.I. Kemp, Considerations when designing human performance tests in the forensic sciences, Aust. J. Forensic Sci. 50 (2018) 166–182, https://doi.org/10.1080/00450618.2016.1229815.

Chapter 7
Integrated Discussion

The central topic of this thesis is whether non task relevant information biases a forensic odontology identification opinion. Addressing this question required foundational knowledge of the evaluation, judgement, and decision process in forensic odontology, which proved to be an under researched area despite identification being the main activity in forensic odontology. This lack of empirical evidence has led to prevailing assumptions regarding the evaluation of dental evidence, its interpretation, and the assignment of conclusion categories in the identification opinion. Therefore, phase I of the research presented in this thesis addressed the basis of these assumptions, and phase II applied the knowledge gained to the development of the method to address the main research question of whether non task relevant information biases forensic odontology identification.

The results of phase I of the research are presented in Chapters 3 and 4 and phase II in Chapters 5 and 6 as published manuscripts.

The literature scoping review presented in Chapter 3 verifies the fundamental assumption that radiographs are a valuable component of dental records for identification, providing high confidence and certainty to the outcome. While most currently available published works do not meet PCAST's criteria for foundational validation there is consistent evidence that even a single pair of ante and post mortem intraoral dental radiographs afford enough information for evaluating and deciding identification. Therefore, although radiographic comparison is only one part of the whole comparative methodology, it can be used to represent the evaluation and decision process. This understanding was pivotal for the experimental design used later in phase II of the thesis as this feature allowed isolation and testing of the effects of extraneous circumstantial information on the identification opinion.

Chapter 4 tests the assumption that the categories used in identification scales simply represent identification confidence by extending the analysis of one of the existing accuracy validation studies identified in the scoping review. This reference validation study was chosen as in it, identification categories were selected after comparing pairs of radiographs. A subset of these radiographic pairs was used in the main experiment in this thesis. Importantly, although, in this previous work the primary validating outcome of interest was a binary decision of match/non match, the identification scales that were provided as adjunct choices allowed interpretation and application of the identification categories. The results of my study support the assumed correlation between self rated confidence and the categories of identification used in forensic odontology opinions. This is important to this current research as context bias has been shown to influence decision confidence [1–4] therefore, a change in category decision in the presence of contextual suggestion would infer a cognitive biasing effect. Although in Chapter 4 a correlation between confidence and category was established it remained uncertain whether category choice reflected only probabilistic estimates or also metacognitive confidence. Additionally, the connotation and number of available categories in the scale seemed to influence the fluidity of interpretation and choice, implying that terms in different scales cannot be compared directly. For example, the category "Possible" was sometimes substituted with "Insufficient information" when using scales on which this was an option. In this instance, "Possible" appears to be used to convey sub exclusion judgement, which contravenes the intended use of this category as stated in the guidelines. These findings had a bearing on the major experiment in phase II of this research. To reduce ambiguity, in the experiment in phase II, the options available to participants were restricted to Identified, Probable, Possible, and Exclude. The participants were then asked if "insufficient" would have been their choice if it had been available. Additionally, the noted conceptual ambiguity of self rated confidence prompted the collection of both probabilistic judgement (JOM, judgement of match score) and separate self rated metacognitive confidence scores in the main experiment.

Chapters 5 and 6, present the results of the experiment central to the phase II research. Chapter 5 demonstrates that strong supportive context information increased probabilistic weight estimates (JOM scores) and the odds of a correct decision when comparing radiograph pairs to make a binary choice. Interestingly, there was no significant difference in performance between the forensic odontologists and general dentists, although the small sample size may run the risk of introducing a type 2 error. However, when it came to assigning identification categories, a significant difference was noted between the two participant groups, as reported in Chapter 6. For the forensic odontologists, a strong supportive context increased the odds of choosing "Identified" compared to a strong contradictory context, but this was not seen in the general dentists. This was despite the overall lower frequency of the use of the category "Identified" by forensic odontologists. Forensic odontologists were found to be more reserved in their use of this category and unlike the general dentists, the option of "insufficient information" was never selected when "Identified" was initially chosen. Furthermore, despite being more conservative, context appeared to influence the choice of categories for forensic odontologists more systematically than for general dentists.

This body of work provides evidence suggesting that non task relevant contextual information exerts cognitive effects and influences both the process and the outcome of an expert opinion. However, dualistic interpretations can be applied to these results and their possible implications. First, given the magnitude of the effect found, the limited sample size and ecological applicability of the main experiment, the effect of context has not been proven irrevocably, and hence, recommending management of context information may be premature. Second, even if the result is generalisable, the magnitude and direction of bias do not seem to associate with a significant increase in error rate; on the contrary, it increased the accuracy. Therefore, it may be unnecessary to manage the context and contextual information associated with a case requiring forensic dental identification. However, both may not be cogent arguments when non task relevant contextual information is examined from the perspective of its role and effect in an expert opinion, its contribution to error and error rate in a decision, and the quality of the decision in an expert opinion.

The concept of contextual information is central to clarifying the role, principle, and value of forensic science in the consideration of the ultimate issue (e.g., identification, guilt, culpability etc) in the judiciary system [5]. The emphasis on "concept" is critical because contextual information presents in different forms [6,7]. All conscious and unconscious sensory and perceptual inputs are information, even contextual cues from professional or social interaction or the requirement of the task including non blinded peer review [6,7]. Although not usually recognised as contextual information, any of this information can potentially set expectations and induce contextual bias. Furthermore, it is the nature of the task that specifically defines what information is relevant to the task rather than the domain or discipline [5]. Therefore, context information conceptually refers to all forms of non task relevant information, as reflected in the introduction and Chapter 2 of this thesis. Having a clear and uniform concept is important in determining which information is relevant and which is not, as the use of non task relevant information can have implications for the concept and function of an expert judgement [8–10].

The role of a forensic expert is to provide an independent scientific opinion based on the evidence and relevant information [8,9]. The ultimate issue of identification of a deceased requires consideration of all sources of evidence both circumstantial and scientific by the legal decision maker [11–14]. This doctrine requires that all evidence including scientific opinion is evaluated independently of other sources of evidence [8,15]. Non independence undermines the value because it violates the underlying scientific statistical Bayesian principle, and actually “re counts” the same evidence [8,9,11]. While it may seem clear why non task relevant information should not be used, the need for contextual information is entrenched in the human psychological sense of information and decisional confidence [16–19]. Contextual information may appear to increase decisional accuracy, as was seen in Chapters 5 and 6, which may raise concerns regarding censorship and may advocate for the use of such information. However, this logic violates the principle and value of the independent perspective of forensic science evidence and is a case of overstating the weight of the evidence. Based on these arguments, context information should be managed regardless of whether it is empirically proven to be biasing or not. Such an approach has been adopted by the Netherlands Forensic Institute’s firearm team [20] where despite failing to find empirical proof of contextual bias effect [21] they have advocated and initiated routine management of contextual information.

One of the greatest concerns about contextual bias is the potential for error, ergo a miscarriage of justice, that may result [22]. The current legal emphasis on error rates as part of reliability may demand the risk of error from contextual bias be quantified in the future [23–25]. Quantification and proof of the role of contextual bias in the errors of real world cases are difficult to establish [26,27]. Except for the case of Brandon Mayfield *See footnote[2] where the cognitive biasing effect of "circular reasoning" was identified as a contributory error [28], few cases of miscarriage of justice can directly be attributed to the effect of cognitive bias. Most reported cases of miscarriage of justice or wrongful conviction are attributed to "human error" [29], which implies a lack of intention or misconduct but is also unavoidable. When human error emerged as a scientific concept and a field of safety science, it was believed that although inevitable it can be mitigated [30,31]. In the forensic literature, thematic focus, or keywords such as "human factor" or "human element" are associated with cognitive or context bias [32–35]. These errors may result from and reside in a system of poor management of human factors such as the organisation's hierarchical social structures [31]. Significantly, contextual bias can cascade down or even snowball along the whole evidence trail unless checks are in place to prevent such errors [36].

Existing models and frameworks for the management of human error in fields such as medicine, and aviation safety [30,37,38] could possibly be adapted and applied to quantify and manage errors associated with human factors in forensic odontology. A similar approach has already been proposed by the fingerprint human factor working group in their positional document [39]where one of the recommendations was the need to shelter the practitioner from contextual information to reduce human errors. One important consideration, however, is that due to the multifaceted and complex nature of context information, experimentally derived context bias related error rates may not be directly applicable to casework [25]. Rather, these error rates may be the "tip of the iceberg" that heralds the need to address or include contextual factors in foundational error estimates

[2] Brandon Mayfield, an Oregon Muslim attorney, was arrested by the FBI in relation to a terrorist attack on a train in Madrid, Spain, in 2004. The FBI had identified the prints found on the bomb detonator as belonging to Mayfield. After the arrest, the Spanish National Police informed the FBI that they had identified an Algerian national as the source of the fingerprint. The Mayfield arrest was overturned after the FBI laboratory examined the Algerian fingerprint and Mayfield was released from custody [28].

for a method and are often underemphasised as noted in the discussion in Chapter 3 and by other authors[40,41].

A binary concept of error rate or complementary accuracy is a common notion of error adopted in research [42]. Yet, inter , and intra rater variability can be another concept of error, which may have been underemphasised in cognitive forensic research. Context information may increase variability between and within rater because the resulting idiosyncratic cognitive effect depends on the interaction between information and individual cognitive style and traits [43–47]. Furthermore, even if the biasing effect is the same, the behaviour and choice outcomes can differ between individuals [47]. Kahneman et al. 2021 [48] conceptualised this as "systematic noise" and believed that variability is as important as bias in a system that is supposed to provide a uniform result when given the same information. Forensic literature has recently conceptualised this as "reliability"[49,50] differentiating this from the more researched "bias" or "biasability" [3,51–54]. Traditional disciplines such as forensic odontology rely on the judgement of the practitioner in their comparative methods; therefore, when the variability in decision outcome among practitioners is viewed collectively, it may be considered an error in the method and system. Inter operator variability in decisional outcomes is to be expected in forensic odontology identification given the reliance on experience and tacit knowledge, if however, it is due to non task relevant information this may be undesirable or problematic [48]. Variability can be reduced through "decisional hygiene"; if contextual information is managed even when the context effects have not affected accuracy, because it pertains to the quality of the decision and, therefore, of the method [48].

Quality in a decision implies that the correct outcome is reached for the correct reason(s), which is critical in forensic science practice [32,46]. As noted in Chapters 5 and 6 contextual information can increase accuracy and confidence e.g., "Probable" can be reassigned as "Identified" in the presence of supportive but non task relevant information. Even though the choice is "more accurate", it cannot be said to be a good one because contextual bias has given the evidence extra weight [32,46]. Improper allocation of evidential weight may mislead the factfinder, leading to a miscarriage of justice or even an erroneous conviction [27]. Although it may be obvious how overstating the weight is potentially a serious issue, underestimating the weight of evidence may also be of concern. It can be difficult to assess the quality of a decision, particularly for the intermediate categories in the identification scales [55]. Firstly, because these terms lack binary clarity, they are never wholly inaccurate. When the evidence is truly limited categories, such as "insufficient evidence," "Possible," and "Probable" are appropriate. However, these categories which may suggest "cannot tell" or "identification and exclusion are both possible" can also be used incorrectly as they can be perceived as less committed and therefore default and safe decisions. Secondly, the choice may also be influenced by context, including both the broader environmental context and the microcosm of scale design, as discussed in Chapter 5. As seen in Chapter 4, the category "insufficient evidence" which was presented just before "exclude" was interpreted as "insufficient for exclusion" rather than the actual defined status of "insufficient information for interpretation". Thus, rating scales can contribute to non consensus even when weight evaluation is similar. In the absence of the rationale informing the evaluation process and verifiable ground truthing, assessing the quality or correctness of a decision will have to rely on agreement or consensus. Even though consensus does not equate to accuracy, which is one part of quality, consensus is necessary to determine quality when there is no verifiable perfect assurance of accuracy. Confusion and debate are inevitable results of this circular paradox of determining quality and correctness in the absence of a fundamental truth [41,56–58] *See footnote[3]. The lack of consensus can be an issue, as Cole 2016 [59]noted in relation to the McKie case: "... in fact, we have no way of knowing that Shirley McKie did not make the print in question, other than through the consensus judgment of latent print examiners. In McKie (unusually), there is not even a complete consensus." Additionally, the absence of consensus among expert opinions can result in confusion for factfinders, and in some cases the failure to consult expert opinion on its significance has resulted in mistrial [27]. Consensus on management and the definition of non task relevant information may improve the consensus rate in opinion as the decision will be based on uniform information.

[3] These citations pertain to published works on the issues of concept of consensus versus accuracy. Dror et al 2018[56] comments on confusion of concept of consensus for accuracy in study by Oliver 2017[57]. While Weller and Morris 2020 [58] comment on the use of consensus as a substitute for correctness for evaluating the categories such as "inconclusive" by Dror and Scurich 2020 [41].

Standardisation of practice for quality management may help facilitate the progress towards reaching a consensus about the management of non task relevant information. Ongoing work by the International Organization for Standardisation committee ISO/TC 272 [60] is proof of the recognition of the need for standardisation in forensic science. Although management of non task relevant information in forensic odontology is not addressed specifically, the precedent for conformity and uniformity in forensic practice may support the development of consensus. Still, such standardisation is inadequate for assessing the quality of decisions. Currently, evidence based protocols for assessing the quality of decisions and consensus resolution in forensic odontology are currently underexplored.

Another pertinent topic may be whether accurate consensus can be attained in such contexts where judgement and decision making depend on experience and implicit learning, and if so, whether calibration can be improved through training. Research posits that experts in such domains possess an implicit sense of the frequency of occurrence of the comparative features [61,62]. This characteristic is an adaptation of the innate human ability for statistical learning of the natural environment. In domains without concrete statistical references, such implicit learning is often valid, and close to the actual ground truth and pooled estimates of individual scores can surpass individual performance scores. It is therefore posited that statistical learning ability can be trained and refined to improve calibration and accuracy [61].

Considerations and limitations of the research design

The sample

Validation of expertise requires proof that experts are capable of the claims they make [63]. This often takes the form of demonstrating superior performance when compared to non experts or laypersons. Defining the "layperson" is an important consideration in determining the appropriate comparison group. For example, although skills are often believed to be non transferable [64] certain professional groups, such as dental assistants, radiologists, or anthropologists, have different levels of skill in reading radiographs both between themselves and when compared to the general population. As such, including these professionals collectively as a layperson group will confound the results. However, since interpreting and comparing radiographs requires knowledge and skill, a lay comparison group without any knowledge may not be a meaningful comparison either. The specific skills required for dental radiograph evaluation and matching are implicitly acquired in clinical practice and therefore are common to both general dentists and forensic odontologists, therefore, general dentists are a more appropriate and meaningful non expert comparison group. However, because a formal training pathway and recognition process may not always be available or applicable in forensic odontology, the qualification and quantification of what constitutes a specialist forensic odontologist are variable. The variation in the quantification of the forensic odontology expert group is not unique to the main experiment in this research and was also observed in multiple studies reviewed in Chapter 3. Similarly, for the non expert group used in this research, the general dentist participants also had varying levels of education in and exposure to forensic odontology. High variability meant the need for large samples to allow trends or associations to be detected and verified. The small sample size of this study limited such modelling and overall generalisability.

The small sample size is a known corollary of the small number of forensic odontology specialists worldwide. A limited sample size is a prevalent methodological problem in forensic science cognitive research generally. As pointed out by Kerstholt et al. [65]in their discussion about the lack of evidence of contextual bias in their firearm analysis experiment, which also used contextual information as stimuli, small sample size does increase the possibility of type 2 statistical error where the lack of differences or bias is incorrectly accepted.

The current profusion of web based experiments and surveys may have additionally contributed to the lower than anticipated participation rate because the intended experimental group may have grown weary of participating in surveys. In this research, the dropout rate of participation was 60%, despite efforts in the research design to avoid boredom and fatigue by balancing the ideal and required number of tasks and survey length. The online medium did allow for international participation which is important for a global perspective and because of the increased possibility of international collaboration in large scale disasters e.g., the Indian Ocean tsunami mass disaster in 2004 [66,67].

Vignettes contextual information

The challenges of using vignettes to deliver contextual information were discussed in Chapter 5 and are particularly relevant to situations involving weak suggestions. The "push pull factor" required was posited to have resulted in the overall lack of systematic trends compared to the more definitive and imperative suggestions. The decision to use vignettes was based on their established validity [68–70]. Vignettes allow controlled manipulation of the experimental variables and isolation and restriction of information to non dental information. It also allowed non test vignettes with dental information to be inserted into the narrative to mask the actual purpose of the experiment. Despite these advantages of vignettes, it was difficult to control and monitor how well the participants read the vignettes, although the recorded time gave some indication. Making the participant actively reveal and progress the vignettes through to completion at least ensured compulsory contact, even if only skim reading was done. In the initial design, these vignettes were planned to allow for revisits during radiograph comparison. It would have been interesting to see if difficult decisions triggered more reliance on the contextual information by the number and duration of revisits, however, technical complexity restricted this initial plan in the experimental design.

General criticism of cognitive bias studies.

A major criticism of judgement and decision laboratory experiments is the deviation from actual practice workflows and environments [71]. While not disclosing the specific purpose is possible, it is impossible to hide that the participant is engaged in an experiment, so the Hawthorn effect or behavioural adjustments when one is observed are always a possibility. Researchers who study contextual information effects also face the challenge of maintaining participant naivety because of the increased awareness of cognitive bias and context effects resulting from an increase in publications. These factors restrict the external validity of laboratory based experiments, even when the sample size is large enough for confident generalisation of the results. Covert testing with test cases inserted into a routine work environment is believed by many to be the gold standard [42,72] but it is difficult to control contextual information in the routine work environment which can then become a confounding factor. Furthermore, this kind of testing is very difficult in most forensic odontology work environments where the same practitioner is often responsible for all parts of the identification process including postmortem data collection and, in some cases, the sourcing of the antemortem data. At the commencement of this project, the state of knowledge about the cognitive aspects of the forensic odontology identification process also made this type of testing non viable and premature.

The lack of established knowledge in the judgement and decision process when making an identification in forensic odontology also contributes to the uncertainty of whether the category perception phenomenon [73,74] applies in the evaluation of evidence or if there is an intermediate implicit probabilistic estimate step before categorical decisions. In the experimental environment, the inclusion of a dependent variable such as “Judgement of Match” (JOM) (See chapter 5), allowed the computation of binary accuracy and a posited measure of probabilistic certainty (See the section on confidence in Chapter 2). Unfortunately, the sample size of this study did not allow for a robust or meaningful correlation of this measure with category decisions. The same applied to the self rated confidence levels correlation, consequently only a very broad based analysis was performed.

In addition to concept consideration of the JOM measure, the lack of a mid point, and hence the forced choice nature of the scale did introduce some metrological effects. Whilst requiring participants to commit the match and non match simplifies accuracy calculation, the trade off was that a one point difference 1 to 1 (one graduation) becomes conceptually different decisions. This may result in a seemingly undesirable serious issue but may in part be an artefactual metrological effect. Midpoints in rating scales also face other issues of ambiguity in concept and interpretation [75], for example, the reasons for choosing may include: “Evidence truly does not allow for a decision”, “Cannot decide”, or “Neutral” the default for experiments and surveys [75], all these reasons are rationally different concepts. The conceptual ambiguity of mid points and the need for binary accuracy calculation were the main reasons for not including a mid point in the experimental design.

References:

[1] Fleisig D. Adding Information May Increase Overconfidence in Accuracy of Knowledge Retrieval. Psychol Rep 2011;108:379–92. https://doi.org/10.2466/04.11.23.PR0.108.2.379 392.

[2] Navajas J, Bahrami B, Latham PE. Post decisional accounts of biases in confidence. Curr Opin Behav Sci 2016;11:55–60. https://doi.org/10.1016/j.cobeha.2016.05.005.

[3] Dror IE, Kukucka J, Kassin SM, Zapf PA. No One is Immune to Contextual Bias—Not Even Forensic Pathologists. J Appl Res Mem Cogn 2018;7:316–7. https://doi.org/10.1016/j.jarmac.2018.03.005.

[4] Kahneman D, Klein G. Conditions for intuitive expertise: a failure to disagree. Am Psychol 2009;64:515–26. https://doi.org/10.1037/a0016755.

[5] National Commission on Forensic Science Human Factors Subcommittee. Ensuring That Forensic Analysis is Based Upon Task Relevant Information. 2015.

[6] Maclean CL, Dror I. A Primer on the Psychology of Cognitive Bias. In: Robertson CT, Kesselheim AS, editors. Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science, and Law. 1st ed., Academic Press; 2016, p. 13–24.

[7] Dror IE. Cognitive and Human Factors in Expert Decision Making: Six Fallacies and the Eight Sources of Bias. Anal Chem 2020;92:7998–8004. https://doi.org/10.1021/acs.analchem.0c00704.

[8] Thompson WC. What role should investigative facts play in the evaluation of scientific evidence? Australian Journal of Forensic Sciences 2011;43:123–34. https://doi.org/10.1080/00450618.2010.541499.

[9] Thompson WC, Vuille J, Biedermann A, Taroni F. The role of prior probability in forensic assessments. Front Genet 2013;4:220. https://doi.org/10.3389/fgene.2013.00220.

[10] Champod C, Evett IW, Jackson G. Establishing the most appropriate databases for addressing source level propositions. Sci Justice 2004;44:153–64. https://doi.org/10.1016/S1355 0306(04)71708 6.

[11] Evett I. The logical foundations of forensic science: towards reliable knowledge. Philosophical Transactions of the Royal Society B: Biological Sciences 2015;370:20140263. https://doi.org/10.1098/rstb.2014.0263.

[12] Cook R, Evett IW, Jackson G, Jones PJ, Lambert J a. A hierarchy of propositions: deciding which level to address in casework. Science & Justice 1998;38:231–9. https://doi.org/10.1016/S1355 0306(98)72117 3.

[13] Jackson G, Jones S, Booth G, Champod C, Evett IW. The nature of forensic science opinion—a possible framework to guide thinking and practice in investigation and in court proceedings. Science & Justice 2006;46:33–44. https://doi.org/10.1016/S1355 0306(06)71565 9.

[14] Evett I w., Jackson G, Lambert JA, McCrossan S. The impact of the principles of evidence interpretation on the structure and content of statements. Science & Justice 2000;40:233–9. https://doi.org/10.1016/S1355 0306(00)71993 9.

[15] Champod C, Evett IW, Jackson G. Establishing the most appropriate databases for addressing source level propositions. Sci Justice 2004;44:153–64. https://doi.org/10.1016/S1355 0306(04)71708 6.

[16] Butt L. The forensic confirmation bias: Problems, perspectives, and proposed solutions Commentary by a forensic examiner. J Appl Res Mem Cogn 2013;2:59–60. https://doi.org/10.1016/j.jarmac.2013.01.012.

[17] Kukucka J. People who live in ivory towers shouldn't throw stones: A refutation of Curley et al. Forensic Sci Int 2020;2:110–3. https://doi.org/10.1016/j.fsisyn.2020.03.001.

[18] Curley LJ, Munro J, Lages M. An inconvenient truth: More rigorous and ecologically valid research is needed to properly understand cognitive bias in forensic decisions. Forensic Sci Int 2020;2:107–9. https://doi.org/10.1016/j.fsisyn.2020.01.004.

[19] Thompson WC. Commentary on: Curley LJ, Munro J, Lages M, MacLean R, Murray J. Assessing cognitive bias in forensic decisions: a review and outlook. J Forensic Sci 2020;65:666–7. https://doi.org/10.1111/1556 4029.14272.

[20] Mattijssen EJAT, Kerkhoff W, Berger CEH, Dror IE, Stoel RD. Implementing context information management in forensic casework: Minimizing contextual bias in firearms examination. Science and Justice 2016;56:113–22. https://doi.org/10.1016/j.scijus.2015.11.004.

[21] Kerkhoff W, Stoel RD, Berger CEH, Mattijssen EJAT, Hermsen R, Smits N, et al. Design and results of an exploratory double blind testing program in firearms examination. Science and Justice 2015;55:514–9. https://doi.org/10.1016/j.scijus.2015.06.007.

[22] Kukucka J, Kassin SM, Zapf PA, Dror IE. Cognitive bias and blindness: A global survey of forensic science examiners. J Appl Res Mem Cogn 2017;6:452–9. https://doi.org/10.1016/j.jarmac.2017.09.001.

[23] President's Council of Advisors on Science and Technology, Report to the president Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature Comparison Methods, Washington DC, 2016. https://obamawhite house.archives.gov/sites/default/f 2016.

[24] Koehler JJ. Fingerprint error rates and proficiency tests: What they are and why they matter. Hastings Law Journal 2008;59:1077–100.

[25] Dror IE. The Error in "Error Rate": Why Error Rates Are So Needed, Yet So Elusive. J Forensic Sci 2020;65:1034–9. https://doi.org/10.1111/1556 4029.14435.

[26] LaPorte GM. Wrongful convictions and DNA exonerations: Understanding the role of forensic science. National Institute of Justice Journal 2018:16.

[27] Smit NM, Morgan RM, Lagnado DA. A systematic analysis of misleading evidence in unsafe rulings in England and Wales. Science and Justice 2018;58:128–37. https://doi.org/10.1016/j.scijus.2017.09.005.

[28] Special Report A Review of the FBI's Handling of the Brandon Mayfield Case Full Report final.pdf. US Department of Justice; Office of the Inspector General 2006. https://oig.justice.gov/sites/default/files/archive/special/s0601/final.pdf (accessed June 29, 2022).

[29] Inquiry on Shirley McKie case blames "human error." BBC News 2011.

[30] Reason J. Human error: models and management. BMJ 2000;320:768–70. https://doi.org/10.1136/bmj.320.7237.768.

[31] Dekker S. The Field Guide to Human Error. The Field Guide to Human Error 2014:248.

[32] Kukucka J, Dror IE. Human factors in forensic science: Psychological causes of bias and error. In: DeMatteo D, Scherr KC, editors. The Oxford Handbook of Psychology and Law, vol. 33. in press, New York:Oxford University Press; 2022, p. 1–12.

[33] Dror IE. Cognitive neuroscience in forensic science: understanding and utilizing the human element. Philosophical Transactions of the Royal Society B: Biological Sciences 2015;370:20140255. https://doi.org/10.1098/rstb.2014.0255.

[34] Found B. Deciphering the human condition: The rise of cognitive forensics. Australian Journal of Forensic Sciences 2015;47:386–401. https://doi.org/10.1080/00450618.2014.965204.

[35] Growns B, Martire KA. Human factors in forensic science: The cognitive mechanisms that underlie forensic feature comparison expertise. Forensic Sci Int 2020;2:148–53. https://doi.org/10.1016/j.fsisyn.2020.05.001.

[36] Dror IE, Morgan RM, Rando C, Nakhaeizadeh S. Letter to the Editor — The Bias Snowball and the Bias Cascade Effects: Two Distinct Biases that May Impact Forensic Decision Making. J Forensic Sci 2017;62:832–3. https://doi.org/10.1111/1556 4029.13496.

[37] Swartz MK. What health care is learning from the aviation industry. Journal of Pediatric Health Care 2015;29:1–2. https://doi.org/10.1016/j.pedhc.2014.09.005.

[38] National Commission on Forensic Science. Root Cause Analysis (RCA) in Forensic Science. vol. 2005. 2016.

[39] Latent print examination and human factors : Gaithersburg, MD: 2012. https://doi.org/10.6028/NIST.IR.7842.

[40] Christensen AM, Crowder CM, Ousley SD, Houck MM. Error and its Meaning in Forensic Science. J Forensic Sci 2014;59:123–6. https://doi.org/10.1111/1556 4029.12275.

[41] Dror IE, Scurich N. (Mis)use of scientific measurements in forensic science. Forensic Sci Int 2020;2:333–8. https://doi.org/10.1016/j.fsisyn.2020.08.006.

[42] Cooper GS, Meterko V. Cognitive bias research in forensic science: A systematic review. Forensic Sci Int 2019;297:35–46. https://doi.org/10.1016/j.forsciint.2019.01.016.

[43] Neuberg SL, Judice TN, West SG. What the need for closure scale measures and what it does not: Toward differentiating among related epistemic motives. J Pers Soc Psychol 1997;72:1396–412. https://doi.org/10.1037/0022 3514.72.6.1396.

[44] Dror IE, Péron AE, Hind S LL, Charlton D. When emotions get the better of us: The effect of contextual top down processing on matching fingerprints. Appl Cogn Psychol 2005;19:799–809. https://doi.org/10.1002/acp.1130.

[45] Ask K, Granhag PA. Motivational sources of confirmation bias in criminal investigations: the need for cognitive closure. Journal of Investigative Psychology and Offender Profiling 2005;2:43–63. https://doi.org/10.1002/jip.19.

[46] Fraser Mackenzie PAF, Bucht RE, Dror IE. Forensic Judgment and Decision Making. In: Zentall TR, Crowley PH, editors. Comparative Decision Making, Oxford University Press; 2013, p. 385–415. https://doi.org/10.1093/acprof:oso/9780199856800.003.0038.

[47] Baron J. Thinking and Deciding. 4th ed. Cambridge University Press; 2008.

[48] Kahneman D, Sibony O, Sunstein CR. Noise: A flaw in human judgment. First edit. New York, NY: Little, Brown Spark; 2021.

[49] Dror IE, Scherr KC, Mohammed LA, MacLean CL, Cunningham L. Biasability and reliability of expert forensic document examiners. Forensic Sci Int 2021;318:110610. https://doi.org/10.1016/j.forsciint.2020.110610.

[50] Hartley S, Winburn AP, Dror IE. Metric forensic anthropology decisions: Reliability and biasability of sectioning point based sex estimates. J Forensic Sci 2022;67. https://doi.org/10.1111/1556 4029.14931.

[51] Miller LS. Bias among forensic document examiners: Still a need for procedural changes. Journal of Police Science and Administration 1984;12:407–11. https://doi.org/10.1080/00450618.2013.797026.

[52] Miller Larry S. Procedural Bias in Forensic Science Examinations of Human Hair. Law Hum Behav 1987;11:127–63.

[53] Nakhaeizadeh S, Dror IE, Morgan RM. Cognitive bias in forensic anthropology: visual assessment of skeletal remains is susceptible to confirmation bias. Sci Justice 2014;54:208–14. https://doi.org/10.1016/j.scijus.2013.11.003.

[54] Stevenage S v., Bennett A. A biased opinion: Demonstration of cognitive bias on a fingerprint matching task through knowledge of DNA test results. Forensic Sci Int 2017;276:93–106. https://doi.org/10.1016/j.forsciint.2017.04.009.

[55] Dror IE, Langenburg G. “Cannot Decide”: The Fine Line Between Appropriate Inconclusive Determinations Versus Unjustifiably Deciding Not To Decide. J Forensic Sci 2019;64:10–5. https://doi.org/10.1111/1556 4029.13854.

[56] Dror IE, Kukucka J, Kassin SM, Zapf PA. When Expert Decision Making Goes Wrong: Consensus, Bias, the Role of Experts, and Accuracy. J Appl Res Mem Cogn 2018;7:162–3. https://doi.org/10.1016/j.jarmac.2018.01.007.

[57] Oliver WR. Effect of History and Context on Forensic Pathologist Interpretation of Photographs of Patterned Injury of the Skin. J Forensic Sci 2017;62:1500–5. https://doi.org/10.1111/1556 4029.13449.

[58] Weller TJ, Morris MD. Commentary on: I. Dror, N Scurich "(Mis)use of scientific measurements in forensic science" Forensic Science International: Synergy 2020 https://doi.org/10.1016/j.fsisyn.2020.08.006. Forensic Sci Int 2020;2:701–2. https://doi.org/10.1016/j.fsisyn.2020.10.004.

[59] Cole SA. More than Zero : Accounting for Error in Latent Fingerprint Identification. vol. 95. 2016.

[60] Standards by ISO/TC 272 Forensic sciences. International Organization for Standardization 2022.

[61] Growns B, Martire KA. Forensic feature comparison expertise: Statistical learning facilitates visual comparison performance. J Exp Psychol Appl 2020;26:493–506. https://doi.org/10.1037/xap0000266.

[62] Martire KA, Growns B, Navarro DJ. What do the experts know? Calibration, precision, and the wisdom of crowds among forensic handwriting experts. Psychon Bull Rev 2018;25:2346–55. https://doi.org/10.3758/s13423 018 1448 3.

[63] Martire KA, Edmond G. Rethinking expert opinion evidence. vol. 40. 2017.

[64] Edmond G, Towler A, Growns B, Ribeiro G, Found B, White D, et al. Thinking forensics: Cognitive science for forensic practitioners. Science & Justice 2017;57:144–54. https://doi.org/10.1016/j.scijus.2016.11.005.

[65] Kerstholt J, Eikelboom A, Dijkman T, Stoel R, Hermsen R, van Leuven B. Does suggestive information cause a confirmation bias in bullet comparisons? Forensic Sci Int 2010;198:138–42. https://doi.org/10.1016/j.forsciint.2010.02.007.

[66] Schuller Götzburg P, Suchanek J. Forensic odontologists successfully identify tsunami victims in Phuket, Thailand. Forensic Sci Int 2007;171:204–7. https://doi.org/10.1016/j.forsciint.2006.08.013.

[67] Kieser JA, Laing W, Herbison P. Lessons Learned from Large scale Comparative Dental Analysis Following the South Asian Tsunami of 2004. J Forensic Sci 2006;51:109–12. https://doi.org/10.1111/j.1556 4029.2005.00012.x.

[68] Kim J. Scenarios in information seeking and information retrieval research: A methodological application and discussion. Libr Inf Sci Res 2012;34:300–7. https://doi.org/10.1016/j.lisr.2012.04.002.

[69] Spalding NJ, Phillips T. Exploring the use of vignettes: From validity to trustworthiness. Qual Health Res 2007;17:954–62. https://doi.org/10.1177/1049732307306187.

[70] Evans SC, Roberts MC, Keeley JW, Blossom JB, Amaro CM, Garcia AM, et al. Vignette methodologies for studying clinicians' decision making: Validity, utility, and application in ICD 11 field studies. International Journal of Clinical and Health Psychology 2015;15:160–70. https://doi.org/10.1016/j.ijchp.2014.12.001.

[71] Norman G. The Bias in researching cognitive bias. Advances in Health Sciences Education 2014;19:291–5. https://doi.org/10.1007/s10459 014 9517 5.

[72] Towler A, White D, Ballantyne K, Searston RA, Martire KA, Kemp RI. Are forensic scientists experts? J Appl Res Mem Cogn 2018;7:199–208. https://doi.org/10.1016/j.jarmac.2018.03.010.

[73] Goldstone RL, Hendrickson AT. Categorical perception. WIREs Cognitive Science 2010;1:69–78. https://doi.org/10.1002/wcs.26.

[74] Harnad S. Categorical Perception. Encyclopedia of Cognitive Science, Nature Publishing Group: Macmillan; 2003.

[75] DeMars CE, Erwin TD. Neutral or Unsure: Is there a Difference? 2005:1–12.

Chapter 8
Conclusion

This thesis concludes by summarising the key findings of the research and discussing the implications for future practice and research.

This thesis examined whether non task relevant contextual information impacted the judgement and decision making in forensic odontology identification opinions. The results indicate that non task relevant contextual suggestions did influence decision making, confidence, and conclusion. The extent and degree of this effect depended on whether the participant was a forensic odontologist or a general dentist, whether the radiograph pair was an actual match or non match, and whether the contradictory or supportive contextual suggestion was strong or weak. Strong supportive contexts increased the base match and non match probabilistic estimates and accuracy decision rates, especially for the non match base decisions for both the forensic odontology and dentist participants. However, there was evidence that strong supportive contextual information increased the selection of more definitive INTERPOL identification categories (Exclude and Identified) for forensic odontologists more than for general dentists.

These results suggest, therefore, that strong non task relevant information affected the implicit judgement confidence for both groups of participants, but the influence on the explicated INTERPOL categorical decision was mainly seen in the forensic odontologists. However, unlike in the general dentist participants, the definitive category "Identified" was used reservedly and not associated with an increase in incorrect decisions with contradictory contextual influences. This suggests and supports that forensic odontologists better appreciate the legal implications and costs of errors associated with the different categories of identification.

The findings impel and herald the need for management of non task relevant information. Although the research's sample size and in vitro nature may limit generalisability, there is evidence of the influence on the decision process. This evidence of effect, together with the tenet of expert opinion discussed in the preceding section, argues for managing non task relevant contextual information. Sheltering the practitioner from context information is important, as context bias is unconscious and, therefore, cannot be voluntarily “censored out” once one is exposed to it. However, in order to allow management of non task relevant information, what constitutes relevant information needs to be agreed upon and explicated. A reformative epistemological approach to forensic odontology identification decision and opinion may be required to decipher relevant from non relevant information. This may involve examining the theory of knowledge and the basis of an identification decision and deliberating the logic and limits of a forensic odontology opinion. An example may be the examination of whether the Bayesian approach is implicitly and intuitively applied in an identification decision, and what factors are considered in the deliberation process. Also, how the logic and rationale of the process could be best captured and expressed in the report. In addition to consensus on non relevant information, quantifying the evidentiary value of the different information in antemortem records and standardising the sequence of exposure to lead in the workflow may be important. This is because the sequence of exposure to information has been posited to influence judgement and decision making. An example of this is evaluating antemortem information before evaluating postmortem information which can encourage circular reasoning or confirmatory bias because the postmortem information will then be interpreted through the context of the antemortem information. This enhanced uniformity and management of relevant and non relevant information may reduce unwanted inter rater variation and improve consensus in quality assessment and outcomes. Such standardisation will also naturally evolve to facilitate the advancement of consensus on the quality of decisions, which will, in turn, improve and advance the peer review process.

Through the research journey of this work, it was observed that the cognitive aspect of forensic odontology decisions had been understudied. Factors such as the metrological aspect of concluding categorical and human performance considerations during validation studies have not received much research attention. The number, ranking, and nomenclature of categories influence the cognitive scaling and mapping of the judged weight of the evidence, as was verified. The exact terminology used in different scales has different weights and meanings when ranked and positioned differently. Therefore, regardless of the type of scale used, the full range of the scale and guidelines should be made available to the end decision makers to improve the alignment of the intended level of identification.

The current forensic climate will require validation of the scientific foundation of method and technology, especially in the new modalities used to aid identification. For example, antemortem cone beam, postmortem CT scans, and the use of 'selfie' self acquired digital images. These imaging techniques may present different challenges to conventional radiographic imaging, the comparative features of interest may differ in the level of ambiguity, and this may mean a different level of susceptibility to context effect because context effect increases with increased ambiguity in the evidence. Validation of these new technologies should consider human factors and contextual bias in the research designs.

Although it is not possible to directly generalise and quantify the magnitude of contextual information's influence on forensic odontology identification from this research, this thesis provides evidence that this influence does exist. No matter how strong the proof is or how extensive the context effect is, the principle of forensic science evidence suggests against using non task relevant information. Hence, in the interests of discipline credibility, admissibility of evidence and legal and community trustworthiness, a two pronged approach to managing non task and task relevant information should be the focus of future research.

Appendices

Ethics for research

HUMAN RESEARCH ETHICS COMMITTEE

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Notification of Expedited Approval

To Chief Investigator or Project Supervisor:	**Professor Jane Taylor**
Cc Co-investigators / Research Students:	**Dr Denice Higgins** **Dr Sher-Lin Chiam** **Doctor Mark Page**
Re Protocol:	**The impact of contextual information on judgement and decision making in forensic dental identification**
Date:	**12-Apr-2018**
Reference No:	**H-2018-0058**
Date of Initial Approval:	**12-Apr-2018**

Thank you for your **Response to Conditional Approval (minor amendments)** submission to the Human Research Ethics Committee (HREC) seeking approval in relation to the above protocol.

Your submission was considered under **Expedited** review by the Ethics Administrator.

I am pleased to advise that the decision on your submission is **Approved** effective **12-Apr-2018**.

In approving this protocol, the Human Research Ethics Committee (HREC) is of the opinion that the project complies with the provisions contained in the National Statement on Ethical Conduct in Human Research, 2007, and the requirements within this University relating to human research.

Approval will remain valid subject to the submission, and satisfactory assessment, of annual progress reports. *If the approval of an External HREC has been "noted" the approval period is as determined by that HREC.*

The full Committee will be asked to ratify this decision at its next scheduled meeting. A formal *Certificate of Approval* will be available upon request. Your approval number is **H-2018-0058**.

If the research requires the use of an Information Statement, ensure this number is inserted at the relevant point in the Complaints paragraph prior to distribution to potential participants You may then proceed with the research.

Conditions of Approval

This approval has been granted subject to you complying with the requirements for *Monitoring of Progress*, *Reporting of Adverse Events*, and *Variations to the Approved Protocol* as detailed below.

PLEASE NOTE:
In the case where the HREC has "noted" the approval of an External HREC, progress reports and reports of adverse events are to be submitted to the External HREC only. In the case of Variations to the approved protocol, or a Renewal of approval, you will apply to the External HREC for approval in the first instance and then Register that approval with the University's HREC.

- ***Monitoring of Progress***

Other than above, the University is obliged to monitor the progress of research projects involving human participants to ensure that they are conducted according to the protocol as approved by the HREC. A progress report is required on an annual basis. Continuation of your HREC approval for this project is conditional upon receipt, and satisfactory assessment, of annual progress reports. You will be advised when a report is due.

- ***Reporting of Adverse Events***

1. It is the responsibility of the person **first named on this Approval Advice** to report adverse events.
2. Adverse events, however minor, must be recorded by the investigator as observed by the investigator or as volunteered by a participant in the research. Full details are to be documented, whether or not the investigator, or his/her deputies, consider the event to be related to the research substance or procedure.
3. Serious or unforeseen adverse events that occur during the research or within six (6) months of completion of the research, must be reported by the person first named on the Approval Advice to the (HREC) by way of the Adverse Event Report form (via RIMS at https://rims.newcastle.edu.au/login.asp) within 72 hours of the occurrence of the event or the investigator receiving advice of the event.
4. Serious adverse events are defined as:
    - Causing death, life threatening or serious disability.
    - Causing or prolonging hospitalisation.
    - Overdoses, cancers, congenital abnormalities, tissue damage, whether or not they are judged to be caused by the investigational agent or procedure.
    - Causing psycho-social and/or financial harm. This covers everything from perceived invasion of privacy, breach of confidentiality, or the diminution of social reputation, to the creation of psychological fears and trauma.
    - Any other event which might affect the continued ethical acceptability of the project.
5. Reports of adverse events must include:
    - Participant's study identification number;
    - date of birth;
    - date of entry into the study;
    - treatment arm (if applicable);
    - date of event;
    - details of event;
    - the investigator's opinion as to whether the event is related to the research procedures; and
    - action taken in response to the event.
6. Adverse events which do not fall within the definition of serious or unexpected, including those reported from other sites involved in the research, are to be reported in detail at the time of the annual progress report to the HREC.

- ***Variations to approved protocol***

If you wish to change, or deviate from, the approved protocol, you will need to submit an *Application for Variation to Approved Human Research* (via RIMS at https://rims.newcastle.edu.au/login.asp). Variations may include, but are not limited to, changes or additions to investigators, study design, study population, number of participants, methods of recruitment, or participant information/consent documentation. **Variations must be approved by the (HREC) before they are implemented** except when Registering an approval of a variation from an external HREC which has been designated the lead HREC, in which case you may proceed as soon as you receive an acknowledgement of your Registration.

Linkage of ethics approval to a new Grant

HREC approvals cannot be assigned to a new grant or award (ie those that were not identified on the application for ethics approval) without confirmation of the approval from the Human Research Ethics Officer on behalf of the HREC.

Best wishes for a successful project.

Associate Professor Helen Warren-Forward
Chair, Human Research Ethics Committee

For communications and enquiries:
Human Research Ethics Administration

Research & Innovation Services
Research Integrity Unit
The University of Newcastle
Callaghan NSW 2308
T +61 2 492 17894
Human-Ethics@newcastle.edu.au

RIMS website - https://RIMS.newcastle.edu.au/login.asp

Linked University of Newcastle administered funding:

Funding body	Funding project title	First named investigator	Grant Ref

Information statement for the experiment

INFORMATION STATEMENT

Professor Jane Taylor OAM
BDS, BScDent(Hons), MScDent, Grad Cert PTT, PhD
Discipline of Oral Health
The University of Newcastle
Ourimbah NSW 2258
Phone: (02) 4349 4545 Fax: (02) 4349 4567
Jane.Taylor@newcastle.edu.au

Information Statement for the Research Project:
"Judgement and decision making in forensic dental identification"
Document Version 1; dated 29/11/2017

The Research Team
Professor Jane Taylor. OAM. BDS, MScDent, PhD, FICD
Dr Denice Higgins. BDS, GradDipForOdont, PhD
Dr Mark Page. BDSc(Hons), GradDipClinDent, GCEd, MHAP, PhD,
Dr Sher-lin Chiam. BDS(Sing.) ADC cert, GDipForOdont, FRACDS

You are invited to participate in the research project identified above which is being conducted by Sher-lin Chiam as part of her PhD studies at the University of Newcastle under the supervision of Professor Jane Taylor and Dr Mark Page from the School of health Sciences at the University of Newcastle and Dr Denice Higgins the School of Dentistry at the University of Adelaide.

Why is the research being done?
This study will examine confidence in decision-making. The information collected will aid in refinement of the categorical scale currently used by forensic odontologists for identification.

Who can participate in the research?
We are looking for
(1) dentists with or without formal qualifications in forensic odontology who are currently practicing forensic odontology
(2) dentists who have no practical experience in forensic odontology
(3) lay people who have no dental background or experience

What will you be asked to do?
If you agree to participate, you will be asked to participate in 2 web-based exercises a minimum of 16 weeks apart.

The exercise will contain a number of forensic odontology identification tasks which represent the range of identification cases commonly encountered and will be presented in such a way as to reproduce the working conditions commonly encountered in most forensic odontology units internationally.

These cases are from actual cases obtained internationally and have been totally de-identified.

A case history for each identification case will be provided to aid in your decision when comparing post-mortem and ante mortem dental radiographs.

You will be asked to interpret and compare the radiographs and to indicate your decisions and your confidence in your decisions.

There will also be some general knowledge and general forensic questions. You will be required to answer these questions and state your confidence in your answer.

1

This study will be conducted on-line and access is via a link to the website. You will be given an index number to de-identify you. This index number will also allow you to log back in and continue with the study where you have left off. This means that you can do this study over a few sessions if it is more convenient for you.

At the end of the tasks, there will be a short questionnaire about professional and practice experience and educational history.

If you think it is relevant to you, continuing education points can be obtained by contacting the research manager who will not be one the researchers. A link will be provided on the website for this.

What choice do you have?
Participation in this research is entirely your choice. Those who give their informed consent will be included in the project. Whether or not you decide to participate, your decision will not disadvantage you in any way.

If you do decide to participate, you may withdraw from the project at any time without giving a reason and have the option of withdrawing any data related to previous engagement with the project.

How much time will it take?
Each cycle should take about 45-60 minutes to finish, however, you can complete the task over multiple sittings if this is more convenient. There is no time limit for the completion of each task.

***What are the risks and benefits of participating*?**
There are no risks anticipated from participation. The only graphic content in the case information are the dental radiographs and only essential forensic information is provided.

There will be no direct benefit to you participating, although participation may contribute to continuing professional development.

How will your privacy be protected?
Individual index numbers will be provided to each participant to allow repeated access to the web based tasks. Coding of participant data and results will be via this index number only. Individual participants will not be identified in any publication resulting from this research.

The collected data will be stored securely (on a password protected computer / in a locked filing cabinet) in the Chief Investigator's office, at the Ourimbah Campus of the University of Newcastle. Data will be retained for a minimum of 5 years as per University of Newcastle requirements.

How will the information collected be used?
Results of this study will be presented at the national and international conferences, published in scientific journals and will be published in Sher-lin Chiam's PhD thesis. Individual participants will not be named or identified in any reports arising from the project although individual anonymous responses may be quoted.

Non-identifiable data may be also be shared with other parties to encourage scientific scrutiny, and to contribute to further research and public knowledge, or as required by law.

You can access a summary of the results of the research by contacting Professor Jane Taylor Jane.Taylor@newcastle.edu.au.

***What do you need to do to participate*?**
Please read this Information Statement and be sure you understand it before you consent to participate. If there is anything you do not understand, or have questions, please contact the researcher.
If you would like to participate, please access the website by following this link …………. and complete the consent page to gain access to the test.

Further information

If you would like further information on this research project please contact **Dr Sher-lin Chiam** (Mobile) 0400 125 835, or (Email) Sher-Lin.Chiam@uon.edu.au

Thank you for considering this invitation.

Professor Jane Taylor OAM
Chief Investigator

Complaints about this research
This project has been conditionally approved by the University's Human Research Ethics Committee, Approval No. H-xxx

Should you have concerns about your rights as a participant in this research, or you have a complaint about the manner in which the research is conducted, it may be given to the researcher, or, if an independent person is preferred, to the Human Research Ethics Officer, Research Services, NIER Precinct, The University of Newcastle, University Drive, Callaghan NSW 2308, Australia, telephone (02) 4921 6333, email Human-Ethics@newcastle.edu.au.

Consent for participating in the experiment.

(Presented online before entering the actual web based experiment)

Welcome to the research study!

We would like to invite you to participate in a research project looking into expression of confidence level in forensic odontology identification decision. This research is conducted by researchers at the University of Newcastle, Australia.

Please click the link below to read the information statement about this project.

Click this link: Information Statement_2018.pdf

- I have read and understood the information statement.
- I agree to participate in the above research project and give my consent freely.
- I understand that the project will be conducted as described in the Information Statement, a copy of which I have retained.
- I understand I can withdraw from the project at any time and do not have to give any reason for withdrawing.
- I understand that the provided information will remain confidential to the researchers.
- I have had the opportunity to have questions answered to my satisfaction.
- If you choose to participate, you will be automatically directed to the survey.

I consent to

- read through de identified case information, compare dental radiographs, and indicate my judgement and decision about the identification of the case, as well as the confidence associated with the decision.
- attempt some general knowledge questions and state my confidence in my answer.
- participate in 2 web based exercises a minimum of 16 weeks apart.
- provide my email address so that a unique index number can be given to anonymise my answers.
- receive an invitation to the second survey through the email that I provide.
- answer a short questionnaire about my work and educational background.

☐ **I consent and begin the study.**

☐ **I do not consent; I do not wish to participate.**

Please enter an email to be associated with this survey so that a unique index number can be given to you.

This is important to ensure the anonymity of your answers.
This is also important if CPD (Continuing education development) points are applicable to you, as this will be the email address to which the certificate will be sent.
This will be managed by a project manager who is not a researcher in this project; this is to ensure the anonymity of your answers.

Please enter your email address:
____

Q8 Please re enter and validate your email address:
____

End of Block: Statement and Informed Consent

Instructional video for the experiment

The video provides the step by step instructions for completing the experiment that all participants had to watch before attempting the practice case and proceeding to the actual trial.
The instructional video gives the exact task flow and demonstration of the experiment; therefore, it is a good representation of the actual experiment which consisted of a total of 15 cases: 12 actual experimental and three non experimental cases, respectively. The vignettes in the non experimental cases contained relevant information which consisted of the dates of exposure of the radiographs. Apart from the non experimental cases, these type of vignettes with relevant information were also used for the instructional video and practice case.

The link to the instructions for the experiment:

https://www.youtube.com/watch?v=LyHkBf3Mebg&t=27s