

Constructing digital evidence

A study on how cognitive and human factors affect digital evidence

Nina Sunde

Department of Criminology and Sociology of Law

University of Oslo

June 2022

© **Nina Sunde, 2022**

*Series of dissertations submitted to the
Faculty of Law, University of Oslo
No. 181*

ISSN 1890-2375

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Print production: Graphics Center, University of Oslo.

Summary

The topic of the thesis is the cognitive and human factor's influence on the evidential value of digital traces. Digital evidence is of high importance for solving crime. Therefore, it is essential that digital traces are collected, examined, analysed, and presented in a way that safeguards their evidential value and minimises erroneous or misleading outcomes. A large body of research has been concerned with developing new methods, tools, processes, procedures, and frameworks for handling new technology or novel implementations of technology. In contrast, relatively few empirical studies have examined digital forensics (DF) practice. The thesis contributes to filling this knowledge gap by providing novel insights concerning DF investigative practices during the analysis and presentation stages of the DF process. The thesis draws on theories and research from several scholarly traditions, such as DF, forensic science, police science, and cognitive psychology, as well as social science traditions such as digital criminology and science and technology studies (STS). A mixed-methods approach is applied to explore the research question:

How could a better understanding of the DF practitioners' role in constructing digital evidence within a criminal investigation enable mitigation of errors and safeguard a fair administration of justice?

The thesis is made up of five articles exploring the research question from different perspectives.

The first article aims to bring insights about cognitive bias from the forensic science domain into the DF discipline and discusses their relevance and plausible implications to DF casework. The analysis suggests that cognitive and human factors influence decision-making during the DF process and that there is a risk of bias in all its stages.

The second article applies an experimental design (the DF experiment). The article examines two aspects of DF decision-making: First, whether DF practitioners' decision-making is biased by contextual information, and second, whether those who receive similar information produce consistent results (between-practitioner reliability). The results indicate that the context influenced the number of traces discovered by the DF practitioners and showed low between-practitioner reliability for those receiving similar contexts.

The third article applies a qualitative lens to examine how the low between-practitioner reliability materialises itself in the DF reports and whether and how the trace descriptions influence the evidential value in a legal context. The article demonstrates how the DF practitioners interpret the same traces differently and develops the concept of “evidence elasticity” to describe the interpretative flexibility of digital traces. The article shows how the evidence elasticity of the digital traces enables the construction of substantially different narratives related to the criminal incident and how this sometimes may result in misinformation with the propensity to mislead actors in the criminal justice chain.

The fourth article is based on a survey of the DF practitioners’ accounts of their investigative practice during the DF experiment. The article explores how they handled contextual information, examiner objectivity, and evidence reliability during the analysis of an evidence file. The results show that many started the analysis with a single hypothesis in mind, which introduces a risk of a one-sided investigation. Approximately a third of the DF participants did not apply any techniques to safeguard examiner objectivity or control evidence reliability.

The fifth article examines the DF practitioners’ reporting and documentation practices. It centres on the conclusion types, the content relevant to the evidence value, and the applied (un)certainly expressions. The results were compared to a study of eight forensic science disciplines. The analysis showed that the DF practitioners typically applied categorical conclusions or strength of support conclusion types. They used a plethora of certainty expressions but lacked an explanation of their meaning or reference to an established framework. However, the most critical finding was substantial deficiencies in documentation practices for content essential for enabling audit of the DF investigative process and results, a challenge which also seemed shared with other forensic science disciplines.

Sammendrag

Temaet for avhandlingen er kognitive og menneskelige faktorerers innflytelse på bevisverdien av digitale spor. Digitale bevis er av stor betydning å oppklare kriminalitet. Det er derfor essensielt at digitale spor samles inn, undersøkes, analyseres og presenteres på en måte som ivaretar deres bevisverdi og minimerer feilaktige og villedende resultater. Det er forsket mye på nye metoder, verktøy, prosesser og rammeverk for å håndtere ny teknologi eller nye måter å bruke teknologi på. Det finnes derimot relativt få empiriske studier av digital forensisk¹ (DF) etterforskningspraksis. Avhandlingen bidrar til å fylle dette forskningshullet med ny innsikt om DF etterforskningspraksiser i analyse og presentasjonsfasen av dataetterforskningsprosessen. Avhandlingen bygger på teori og forskning fra flere vitenskapelige tradisjoner, som DF, forensisk vitenskap,² politivitenskap, kognitiv psykologi og samfunnsvitenskapelige tradisjoner som digital kriminologi og vitenskaps- og teknologistudier (STS).

En kombinasjon av ulike forskningsmetoder er brukt for å besvare forskningsspørsmålet:

Hvordan kan en bedre forståelse av DF etterforskerens rolle i konstruksjonen av digitale bevis i en straffesak etterforskning gjøre det mulig å forhindre feil og sikre en rettferdig rettergang?

Avhandlingen består av fem artikler som undersøker forskningsspørsmålet fra ulike perspektiv.

Første artikkel tar sikte på å bringe innsikt om kognitive bias fra andre forensiske disipliner inn i DF disiplinen, og diskuterer deres relevans og mulige konsekvenser for DF etterforskningsarbeid. Analysen indikerer at kognitive og menneskelige faktorer påvirker beslutningstakingen i DF prosessen, og at det er risiko for bias i alle fasene.

Andre artikkel bruker et eksperimentelt design (DF eksperimentet). Artikkelen undersøker to aspekter av DF beslutningstaking: For det første, om DF etterforskernes beslutninger ble påvirket av kontekstuell informasjon, og, for det andre, om de som mottok lik informasjon

¹ På norsk brukes også begreper som digital kriminalteknikk og datatekniske undersøkelser.

² Oversatt fra «forensic science». Begrepet omfatter ulike vitenskapelige undersøkelser som kan ha relevans for retten, som rettspsykiatri, rettsmedisin, rettsstoksikologi, rettsgenetikk. Kriminalteknikk og digital kriminalteknikk regnes som en vesentlig del av dette området.

produserte konsistente resultater (mellom-etterforsker reliabilitet). Resultatene indikerer at konteksten påvirket mengden av spor som DF etterforskerne oppdaget, og viste lav mellom-etterforsker reliabilitet for de som mottok lik kontekst.

Tredje artikkel bruker en kvalitativ tilnærming til å undersøke hvordan den lave mellom-etterforsker reliabiliteten materialiserer seg i DF etterforskernes rapporter, og hvordan beskrivelsene av spor påvirker bevisverdien i en rettslig kontekst. Artikkelen viser hvordan DF etterforskerne tolker de samme sporene forskjellig og utvikler konseptet “beviselastisitet” for å beskrive den fortolkningsmessige fleksibiliteten ved digitale spor. Artikkelen demonstrerer hvordan beviselastisiteten i de digitale sporene muliggjør konstruksjonen av vesentlig forskjellige narrativ om den etterforskede hendelsen, og hvordan dette noen ganger kan resultere i feilinformasjon som kan villede aktørene i straffesakskjeden.

Fjerde artikkel er basert på en undersøkelse av DF etterforskernes betraktninger om sin etterforskningspraksis under DF eksperimentet. Artikkelen utforsker hvordan de håndterte den kontekstuelle informasjonen de mottok, samt hva de gjorde for å ivareta sin objektivitet og for å kontrollere bevisets pålitelighet under analysen av databeslaget. Resultatene viser at mange startet analysen med en enkelt hypotese i tankene. Et slikt utgangspunkt introduserer en risiko for en ensidig undersøkelse. Omtrent en tredel av DF etterforskerne brukte ingen teknikker for å ivareta sin objektivitet eller kontrollere bevisets pålitelighet under analysen.

Femte artikkel undersøker DF etterforskernes rapporterings og dokumentasjonspraksis. Mer spesifikt fokuserer artikkelen på anvendte konklusjonstyper, rapportert informasjon med relevans for bevisets verdi og anvendte begreper for (u)sikkerhet. Resultatene sammenlignes med en studie av åtte andre forensiske disipliner. Analysen viser at DF etterforskerne typisk brukte kategoriske konklusjoner eller ‘graden av støtte’ konklusjoner. De brukte en mengde ulike uttrykk for (u)sikkerhet, men uten å forklare uttrykkenes betydning eller å henvise til et etablert rammeverk for slike uttrykk. Det mest kritiske funnet er imidlertid betydelige mangler i dokumentasjonspraksisen for informasjon som er avgjørende for en kritisk vurdering av hvordan DF prosessen ble gjennomført og dens resultater. Denne utfordringen ser ut til å være gjeldende også for de forensiske disiplinene det ble sammenlignet med.

Acknowledgements

The generous help, inspiration, encouragement, and support from many people have been invaluable and have made my academic journey a wonderful and rewarding learning experience.

I would first like to thank my supervisors, Helene Oppen Ingebrigtsen Gundhus, Johanne Yttri Dahl, Fergus Thomas Toolan, and Itiel E. Dror for your insightful guidance and support. I am thankful for the helpful feedback and insightful comments by Corinna Kruse during the midway assessment seminar and the valuable feedback and advice from Heidi Mork Lomell during the final seminar.

I am grateful to the Norwegian Police University College for funding the research, and the current and former principals Nina Skarpenes, Tor Tanke Holm, Siw Hansen Thokle, Kjell Eirik Mortensen, Ivar Husby, John Ståle Stamnes, and Dag Mørk Sveaas for their support and providing the necessary working conditions for completing the thesis within the planned timeframe. I am thankful to the Department of Criminology and Sociology of Law at the University of Oslo for accepting my project and guiding me steadily through the PhD programme.

I want to thank Torstein Eidet, Jørn Helge Jahren, Torstein Schjerven, Christine Sætre, Nadja Kirchhoff Hestehave, Kristina Kepinska Jakobsen, Inger Marie Sunde, Morten Holmboe, Jon Strype, Gunnar Thomassen, Håvard Aanes, Torbjørn Skardhamar, Camilla Pellegrini Meling, Brita Bjørkelo, Janne H. I. Helgesen, Anita M. Tveit, Rune Olav Andersson, Sølvi-Agnete Olstad, Jon Aga, Inger-Lise Brøste, Øystein Skjønborg, Marthe L. Sakrisvold, Eivind Kolflaath, Lillian Bøylestad, Lene Wachter Lentz, Graeme Horsman, Andreas Røed, Moa Lidén, and members of the NCFI team for their valuable inputs, help, insights, discussions, inspiration, and motivation which helped me to successfully complete my thesis. Thanks to all my colleagues at the Norwegian Police University College and the Norwegian Police Service for their interest in the research project, inspiration, and support. Thanks to the members of the research group Police and Technology for fruitful and inspiring discussions.

I want to thank the DF practitioners participating in the research – which all have invested their valuable time in this research. I am grateful to the anonymous referees for their insightful comments and critiques of the articles.

Eventually, I want to thank my family. I am deeply grateful to my mother, Brit, and father, Roald, for your love, encouragement, and support.

Thanks to my brother, Per Morten, for inspiring discussions and for sharing your valuable knowledge and experience.

Last, but not least – a very special thanks to my dear husband, Henrik, and my beloved sons, Benjamin and Oliver – completing the thesis would not have been possible without your patience, love, and support.

List of tables and figures

Tables:

1. Attrition in the DF experiment according to received context.
2. Overview of the collected material.
3. Overview of the material, type of applied analysis, and associated articles.
4. Overview of themes and codes in Article 3.
5. Code groups of techniques and approaches in Article 4.
6. Conclusion types and values used in the quantitative analysis in Article 5.
7. Content type and values used in the quantitative content analysis in Article 5.

Figures:

1. The DF process, adopted from Flaglien (2018).
2. Average observed traces (of max. 11) per group in Article 2.
3. Average reliability score per group for the observation, interpretation, and conclusion levels in Article 2.
4. An illustration of a combined investigator and theory triangulation in a DF investigation.

Abbreviations

AI	Artificial Intelligence
ANT	Actor-network theory
CAI	Case Assessment and Interpretation
CCTV	Closed-circuit television
DF	Digital forensic(s). The abbreviation is used for both digital forensics (noun) and digital forensic (adjective) in the thesis
DFRWS	Digital Forensic Research Workshop
DNA	Deoxyribonucleic acid
ENFSI	European Network of Forensic Science Institutes
HEP	Hierarchy of Expert Performance
IoT	Internet of Things
ISO/IEC	International Organization for Standardization
LSU	Linear Sequential Unmasking
LSU-E	Linear Sequential Unmasking – Expanded
NIST	National Institute of Standards and Technology
NPCC	The National Police Chiefs’ Council
OSAC	The Organisation of Scientific Area Committees for Forensic Science
STS	Science and technology studies
SWGDE	Scientific Working Group on Digital Evidence
UKAS	United Kingdom Accreditation Service
5WH	What, when, where, why, who and how

Table of contents

Summary	i
Sammendrag (summary in Norwegian)	iii
Acknowledgements	v
List of tables and figures	vii
Abbreviations	viii
PART ONE	5
1. Introduction	6
1.1 Human factors and the construction of digital evidence	6
1.2 Research objective: Unboxing the DF practitioner’s investigative practice	8
1.2.1 Delimitation	14
1.2.2 Central terms	14
1.2.3 Outline of the thesis	17
2. Empirical context	18
2.1 Brief history of DF and its placement in forensic science.....	18
2.1.1 Pre-history.....	18
2.1.2 Infancy	19
2.1.3 Childhood.....	20
2.1.4 Adolescence	21
2.2 Challenges of the DF discipline.....	23
2.2.1 Technical.....	23
2.2.2 Human related.....	25
2.3 DF practitioner conduct	27
2.3.1 Identification, collection, and examination stages.....	28
2.3.2 Analysis stage	29

2.3.3 Presentation stage.....	33
2.3.4 Process dynamics and dependencies.....	35
2.4 Digital evidence agency - increasing volumes and new functions.....	38
2.5 Cognitive and human factors influencing forensic decision-making.....	39
3. Theoretical perspectives on the DF practitioner’s role in the journey from digital trace to evidence	42
3.1 Traceology.....	42
3.1.1 Theorising the trace from a semiotic perspective	43
3.1.2 The reported trace as an actor with agency.....	45
3.2 From trace to storyline – connecting the dots and crafting a scenario	45
3.2.1 DF casework – as scientific inquiry or investigation.....	47
3.2.2 DF casework – as interpretation, narrative construction, and inscription.....	53
3.3 Evidential value of traces	57
3.3.1 Demonstrative tangible evidence	57
3.3.2 Testimonial evidence	59
3.4 Error and uncertainty.....	60
3.4.1 Categorisation of error	60
3.4.2 Cognitive architecture and mechanisms as sources of error.....	62
4. Methods.....	66
4.1 Sample, data collection, and material.....	66
4.1.1 Background studies.....	66
4.1.2 The experiment	67
4.1.3 Strengths and limitations of the research design.....	78
4.2 Analytical procedures	80
4.2.1 Article 2	81
4.2.2 Article 3	84
4.2.3 Article 4	85

4.2.4 Article 5	87
4.3 Research quality and ethical considerations	89
4.3.1 Permissions	89
4.3.2 Anonymisation and handling of the material	90
4.3.3 Research quality	90
4.4 The professional position and scientific worldview	94
5. Summary and integration of results	97
6. Concluding analysis and discussion – implication of findings	104
6.1 A broader understanding of the mutable components of the digital evidence	104
6.2 The DF practitioner’s role as mediator in a technosocial process from trace to evidence	106
6.2.1 DF investigation – science, investigation, or innovation?	107
6.2.2 Mediating relevance, credibility, and evidential value through inscriptions	108
6.2.3 The function of the mutable components for evidence and narrative crafting	111
6.3 Managing unwanted consequences of elastic digital evidence	112
6.3.1 Biasability	113
6.3.2 Bias minimising measures	114
6.3.3 Reliability	117
6.3.4 Noise and variation – friend or foe?	119
6.3.5 Experiment as a “noise” audit	121
7. Summing up the contribution of the thesis	123
7.1 Future research	124
References	126
PART TWO	151
Article 1: Sunde, N., Dror, I. E. (2019). Cognitive and human factors in digital forensics: Problems, challenges, and the way forward. <i>Digital Investigation</i> , 29, 101-108. https://doi.org/10.1016/j.diin.2019.03.011	152

Article 2: Sunde, N., Dror, I. E. (2021). A Hierarchy of Expert Performance (HEP) applied to digital forensics: Reliability and biasability in digital forensics decision making. <i>Forensic Science International: Digital Investigation</i> , 37, 301175. https://doi.org/10.1016/j.fsidi.2021.301175	161
Article 3: Sunde, N. (2022). Unpacking the evidence elasticity of digital traces. (Manuscript submitted for publication).	173
Article 4: Sunde, N. (2021). Strategies for safeguarding examiner objectivity and evidence reliability during digital forensic work. <i>Forensic Science International: Digital Investigation</i> , 40, 301317. https://doi.org/10.1016/j.fsidi.2021.301317	206
Article 5: Sunde, N. (2021). What does a digital forensics opinion look like? A comparative study of digital forensics and forensic science reporting practices. <i>Science & Justice</i> , 61(5) 586-596. https://doi.org/10.1016/j.scijus.2021.06.010	216
Appendices:	228

PART ONE

1. Introduction

1.1 Human factors and the construction of digital evidence

We live in a globalised, digitised, and interconnected society, where technology is woven into almost all aspects of our social relations and activities. Actions, movements, and communication leave digital traces, often without the knowledge or effort of the actor. In fact, *not* leaving traces requires much more effort and sophistication than creating a digital trail. In the context of criminal investigations, digital traces have great value, due to their availability to shed light on the six critical questions (T. Cook, 2016, p. 38) initiating the typical queries that need to be solved to reconstruct an alleged criminal activity, namely, what, where, when, who, why, and how. In 2011, Casey (2011b, p. 3, italics in original) stated, “In this modern age, it is hard to imagine a crime that does not have a *digital dimension*.” The description is also valid today. In a study from 2010, law enforcement agencies in the USA reported that 50% of their cases had a digital component (Gogolin, 2010, p. 4). Alongside the digitalisation trend in society, the proportion has increased, and, according to the Digital Forensic Strategy issued by the UK National Police Chiefs’ Council (2020), over 90% of all recorded crime now has a digital element. A study of 44 homicide cases in the UK showed that digital evidence (including CCTV and phone data) played an integral role in helping to solve the cases and was the most frequent type of forensic evidence for identifying and for charging suspects (Brookman & Jones, 2021, pp. 8-9). Together with physical evidence, digital evidence serves as key anchor points when detectives strive to assemble investigative narratives (Innes et al., 2021, pp. 713-714).

In addition to their availability, digital traces’ popularity as evidence may relate to their aura of objectivity and credibility. According to empirical observations by Innes and colleagues (2021, p. 718), digital evidence is becoming increasingly influential in the investigative sense-making work performed by criminal detectives, due to the credibility it is afforded. Securing digital traces and transforming them into evidence in a forensically sound manner entails securing evidence in the best possible way, preferably by preserving their integrity or by documenting and explaining the relevance and implications of any necessary alterations (Casey, 2011c, p. 233; Casey & Dywalt 2011, p. 398; McKemmish, 2008). Therefore, possessing the necessary expertise is an essential component of a forensically sound process. Further, understanding the traces’ relevance and meaning in the context of the criminal investigation, while considering their limitations and uncertainties, is a complex task that requires knowledge and skills from various domains and disciplines, such as computer

science, digital forensics, law, and the interdisciplinary field of criminal investigation (Sunde, 2017, p. 103). The traces are transformed into evidence in a highly technology- and structure-dependent process. The information processing is performed in a co-construction between the Digital Forensic (DF) practitioner and software, hardware, and tools and is guided by legal rules, principles, standards, and best practice guidelines aimed at safeguarding an outcome – the digital evidence – in compliance with the rule of law.

However, much can go wrong during the DF process, which may lead to loss or alteration of evidence (Casey, 2002; Cohen, 2013, pp. 30-32, 47-48). The Danish telecom case (“Teledata sagen”) showed that systematic errors in evidence handling systems could remain undetected over a long period, due to inadequate quality systems and excessive trust in experts (Lentz & Sunde, 2020; Sorensen, 2019). There are also reports about flawed versions of DF tools commonly used by law enforcement, resulting in erroneous timestamp interpretations (Grut, 2020). The British Post Office scandal showed that the belief that Horizon was a fail-safe system led to what is referred to as the biggest miscarriage of justice in UK history (Flinders, 2021; Virgo, 2021). According to Van Buskirk and Liu (2006), a perception exists among many in the legal community that digital evidence is reliable and correct if accepted and admitted in court. Despite the knowledge about vulnerability to errors, digital evidence is increasingly presented and accepted in courts without scientific validation of the DF methodology or tools (Stoykova, 2021, p. 1). Commonly held techno-fallacies are beliefs that technology is neutral, that *facts* speak for themselves or the belief in a hundred per cent fail-safe systems (Marx & Guzik, 2017, pp. 500-502). These beliefs may originate from assumptions of “mechanical objectivity” (Daston, 1992, p. 599; Daston & Galison, 2007, p. 115), which implies that machines produce richer, better, and truer evidence than a human. Such a perspective is reflected, for example, in rules governing the admissibility of digital evidence in England and Wales, where it is stated as a main rule that computer systems should be considered reliable: “In the absence of evidence to the contrary, the courts will presume that mechanical instruments were in order at the material time” (The Law Commission 1997, para 13.13.).

In contrast to the witness telling their story to the court, the DF practitioners are *the voice* of the digital evidence to the legal decision makers, by either documenting and describing the evidence in reports or presenting the findings orally in court. A substantial body of research from neighbouring forensic science disciplines has shown that the processes involved in producing the evidence are prone to cognitive and human error and may lead to flawed results

(Cooper & Meterko, 2019). However, the research on DF practitioners' roles and practices is sparse, and there is a research gap concerning how human factors influence the value of digital evidence. Therefore, the thesis investigates whether the notion of mechanical objective digital evidence may be justified or whether aspects related to cognitive and human factors are particularly prone to introducing error and uncertainty when digital evidence is constructed. This knowledge is vital because, if an invalid notion of objectivity and credibility leads to inadequate quality control and poor scrutiny by the legal decision makers, there is a risk of the flawed evidence becoming misleading, with miscarriages of justice as a possible consequence.

1.2 Research objective: Unboxing the DF practitioner's investigative practice

Compared to the magnitude of research on street policing and crime control, less attention has been directed towards investigative practices within the police (Holmberg, 2014, p. 172; Reiner, 2010; Stelfox, 2009, pp. 1-3). However, the body of empirical research concerning criminal investigation is growing (e.g., Brodeur, 2010; Dean, 2000; Fahsing, 2016; Gundhus et al., 2022; Hestehave, 2021; Innes, 2003; Rachlew, 2009; Runhovde, 2017). According to Innes et al. (2021, p. 709), the scholarly work on criminal investigation has revolved around the three framings: crime, conduct, and techniques. To this date, the research foundation on DF investigative techniques is extensive, focusing on methods and technology development for securing digital evidence in a constantly changing technological environment. Yet, relatively few studies have examined DF investigative conduct (e.g., Brookman & Jones, 2021; Hansen et al., 2017; Haraldseid, 2021; Jahren, 2020; Rappert et al., 2021; Ward, 2021; Wilson-Kovacs, 2021).

The overarching theme of the thesis is the cognitive and human factor's influence on the evidential value of digital traces. The thesis draws on theories and research from several scholarly traditions, such as DF, forensic science, police science, cognitive psychology, and the social science traditions digital criminology and science and technology studies (STS). Inspired by theory from the STS domain, digital evidence is perceived as an actor with agency (Latour, 1992, p. 227), and the digital evidence's function and social life are central aspects of the research. The main theoretical concepts and perspectives are outlined and debated in section 3. The primary research question is:

How could a better understanding of the DF practitioner's role in constructing digital evidence within a criminal investigation enable mitigation of errors and safeguard the fair administration of justice?

The research question is divided into four sub-questions, annotated with the article(s) in which they are addressed:

- What are the mutable components of the digital evidence? (Article 3)
- What characterises the DF practitioner's practice in the analysis and presentation stages of the DF process? (Articles 2, 3, 4, 5)
- How may the DF practitioner construct or negotiate the mutable components of the digital evidence? (Articles 1, 2, 3, 4 and 5)
- What are the cognitive and human factors that may influence the DF decision-making? (Articles 1 and 2)

The thesis research question is explored through a mixed-methods study that centres mainly on the conduct-framing mentioned above and also, to some extent, on the techniques-framing when examining DF practitioners' practices and decision-making. The data were primarily collected through an experiment ("the DF experiment") involving 53 DF practitioners who analysed the same evidence file and wrote an individual report documenting the analysis and findings, as they would do in a typical DF investigation. Immediately after submitting the result, they completed a survey about how they had conducted the analysis and assessed the findings. Section 4.1 presents and discusses the data collection in more detail.

The primary objective of the thesis is to open the DF process' "black box", a concept described by Latour (1987, p. 4) as "uncertainty, people at work, decisions, competition, controversies are what one gets when making a flashback from certain, cold, unproblematic black boxes of their recent past". Opening the lid of the black box enables insights into the otherwise hidden decisions, procedures and investigative practices during DF casework and the uncertainties concerning the result. Placing normative process descriptions and principles in the background and, instead, exploring *how* they are enacted provides an opportunity to gain insight into the human actor's role in constructing digital evidence in a criminal investigation and how this may influence the result. The version of the DF process referred to in the thesis consists of the identification, collection, examination, analysis, and presentation

stages (Flaglien, 2018). The thesis centres on the analysis and presentation stages that are less technically centred and more investigation oriented. The complete DF process is outlined in section 2.3.

The thesis adds to the field of *police science*, which is defined as “the scientific study of the police and others carrying out policing activities, who they are, their tasks and their societal role, what they do and the effects of it” (Larsson et al., 2014, p. 19, my translation). The thesis is, in a broader sense, a study of investigative practice and knowledge construction by the police, and thus adds to studies on socio-technical knowledge production in police emergency control rooms (Lundgaard, 2019), by police patrols (e.g., Gundhus, 2013; Gundhus et al., 2022; Marciniak, 2021), the use of CCTV footage in policing (Lomell, 2004), and socio-technical processes in crime predictions (e.g., Duarte, 2021; Kaufmann, 2017; Kaufmann et al. 2019; Leese, 2021).

From a *forensic science scholarly perspective*, the thesis extends the empirical knowledge about the vulnerability of the cognitive tasks involved in forensic decision-making, such as “the forensic confirmation bias” (Kassin et al. 2013) and the reliability of forensic science decision-making (see an overview in Cooper & Meterko, 2019). The research regarding and within the DF domain has centred on constructing normative procedures and processes for the technical aspects or physical tasks of DF work; consequently, there is a research gap concerning DF practice and, particularly, the investigative and cognitive tasks involved in DF work. A few scholars from the forensic science or DF domains have researched aspects of DF practices, such as acquisition or preview (Carlton, 2007; Hewling, 2013; J. I. James & Gladyshev, 2013b), quality assurance (Andersson, 2020; Jähren, 2020; Tully et al., 2020) or collaboration with other parties in the criminal investigation (Borhaug, 2019; Hansen et al., 2017; Sunde, 2017). Scholars from the social science domain, such as Fiona Brookman, Helen Jones and Dana Wilson-Kovacs, have conducted ethnographic research involving observation, interviews and document studies. These methods provide valuable insights into the physical conduct and the participant’s accounts of their practice. The limitation is, however, that the *hidden* investigative conduct – such as what they observe, how they interpret, assess, infer, and decide upon their findings – is less available for the researchers’ direct scrutiny. The participants would typically be unaware of the cognitive processes, since they happen unconsciously (Nickerson, 1998, p. 175; Pohl, 2022, p. 7). Interviews thus result in a description of what they think they do, probably combined with what they ought to do.

Although some of the above-mentioned research have touched upon cognitive factors influencing DF investigative tasks, none has specifically targeted the cognitive factors that affect DF work or explored implications such as biased and inconsistent decision-making. Earwaker and colleagues (2020, p. 9) highlighted that such knowledge is vital for improving the transparency and reproducibility of forensic science decision-making. The DF experiment created a situation that disclosed hidden processes in DF decision-making and is currently the first study to explore the issues of bias and reliability in DF investigative work. The DF experiment was designed to achieve the best possible ecological value, which entailed some compromises regarding the experimental conditions. How this challenge was tackled is discussed in detail in sections 4.1.2 and 4.1.3. Although some limitations were identified, the thesis provides valuable insights into how DF practitioners approach their investigative tasks, the perception, interpretation, and decision-making involved during the analysis and presentation stages of the DF process, and how these processes influence the result.

By engaging many DF practitioners to perform a task that a single practitioner would typically undertake, the DF experiment also provides valuable insights into the low reliability /consistency of DF practitioner investigative conduct. Low consistency/reliability in decision-making is referred to as “noise” and generally presented as a problem that should be tackled through preventive measures (Kahneman et al., 2021, p. 370). Based on the empirical findings, the thesis challenges this assumption and aims to nuance the perception of noise, by exploring the function of variance and zooming in on the evidential components or facets the variance relates to, as well as exploring whether there are situations in DF investigations and decision-making where variance may be considered a utility.

As demonstrated in section 2.3, there are many research papers defining how the DF process ought to be performed, but how the stages are performed *in practice* has not been subject to much empirical research. Haraldseid (2021) interviewed six Norwegian criminal detectives considered specialists in this task, about how they performed content analysis (as opposed to technical analysis – see section 2.3.2), and concluded that no standard procedure seemed to be applied to this task. Brookman and colleagues’ research sheds light on the construction process of digital evidence in UK homicide cases, for example, through their descriptions of how CCTV evidence was crafted to underpin the narrative that is to be presented in court (Brookman & Jones, 2021). The thesis devotes particular attention to the analysis and presentation stages, to expand the sparse knowledge about the investigative and reporting/documentation practices

for constructing digital evidence through the DF process. The 53 analysis reports authored by the DF practitioners during the DF experiment, combined with their accounts of how they approached the analysis, provide unique insight into their analysis and reporting practices and elucidate the diversity of practices that might be unknown to the practitioners and difficult to observe through research. The thesis adds insight into how transparency about investigative practices and decisions affects the perceivable evidential value of the result. This knowledge is vital for advancing the theoretical understanding of the hidden processes and decisions involved in DF casework, and for designing adequate measures for auditability, error minimisation and sustaining the necessary quality of the procedures and the results.

From a *criminological perspective*, the thesis provides valuable empirical insights into the theoretical understanding of “technosocial practices” (Stratton et al., 2017, p. 24) within DF work, particularly how digital traces are transformed into evidence by DF practitioners and how the human and non-human entities influence the result. Within the broad field of criminology, the thesis contribution advances the relatively new field of digital criminology, which examines “conceptual, legal, political and cultural framings of crime, formal justice responses and informal citizen-led justice movements in our increasingly connected, global and digital society” (Powell et al. 2018, p. 3). Digital criminology emerged as a reaction to cyber criminology and seeks to extend beyond the traditional topics of cybercrime, policing, and law. Stratton et al. (2017, p. 18) argue that studies of computing and cybercrime have inadequately considered the ongoing technological developments, such as the social web, big data, and the Internet of Things (IoT). It is also argued that the research has been too insular, lacking critical and interdisciplinary engagement with fields such as sociology, computer science, politics, journalism, media, and cultural studies (Powell et al., 2018, p. 3). Digital investigation and evidence are described as a field that needs more attention from researchers: “Importantly, digital evidence is collected and used in ways that require an in-depth understanding of the investigation process. Digital investigations raise new and important questions over how evidence is collected, retained and regulated in relation to privacy and individual liberties” (Powell et al., 2018, p. 30, referring to Kerr, 2005, p. 280).

The thesis expands the empirical knowledge of DF practices investigated by scholars from social science, who have primarily examined practices during the collection and examination stages of the DF process (Rappert et al., 2021; Wilson-Kovacs, 2019) or digital evidence in particular crimes such as homicide (Brookman & Jones, 2021; Brookman et al., 2020a; Innes

et al., 2021) and sexual abuse (Wilson-Kovacs et al., 2021). In particular, the thesis advances the concept of “interpretative flexibility” (Collins, 1981, p. 4; Doherty et al., 2006, p. 569) in a technosocio-legal context, by bridging the elastic/variable components of digital evidence and the legal concepts of evidence reliability, credibility, and inferential/probative force. The research provides insights into how DF practitioners mediate the understanding of what the trace is, its value to the case under investigation, and the plausible scenarios or narratives underpinned by the digital traces. This adds to the knowledge about the role of forensic evidence (including digital evidence) in investigative sense-making and narrative construction, recently explored by Brookman and colleagues through studies of investigative and collaborative practices during British homicide investigations (Brookman & Jones, 2021; Brookman et al., 2020a, 2020b; Innes et al., 2021; Jones et al., 2020).

The layperson’s and legal decision maker’s notion of mechanical objectivity regarding digital evidence appears to persist, despite the substantial body of scholarly literature discussing the systematic and random technical errors that may influence the quality or value of digital evidence. Through its examination of the implications of bias and noise, the thesis expands the knowledge on the non-technical sources of error and uncertainties in digital evidence, as well as on how human and cognitive factors may lead to the construction of misinformation and flawed digital evidence. It sheds light on the limitations and deficiencies in DF practices for maintaining examiner objectivity and evidence reliability during the examination of digital evidence, and on inadequate reporting practices. The thesis also elucidates aspects of digital traces that may nuance the notion of objective and credible evidence due to the demonstration of their elasticity, that is, how differently digital evidence may be constructed in terms of what it is, what it means and its value to the case under investigation. This is demonstrated in quantitative terms through the calculations of between-practitioner reliability and through qualitative analysis of how the variation manifests itself in actual descriptions of traces, trace relationships, and the scenarios underpinned by the traces. The thesis brings about theoretical assumptions of possible mechanisms that contribute to sustaining the notion of objective and reliable evidence, despite the available knowledge about its vulnerability to error, which may be explored in future research.

Based on the empirical foundation, the thesis aims to describe the normative challenges concerning examiner objectivity and evidence credibility. This will form a new basis for

further empirical (practices) and normative (what measures are necessary) discussions and research.

1.2.1 Delimitation

Digital evidence has primarily been associated with cybercrime in criminological research. However, today, digital evidence is obtained in investigations regardless of crime type and severity, due to its prevalence, availability, and ability to inform the core investigative questions. Therefore, the thesis applies a general approach to digital evidence in criminal investigations, in accordance with the current state of the role of such evidence.

The thesis takes a micro level approach and zooms in on a single actor in the criminal justice system: the DF practitioner. Although the interplay with technology, automation, and advanced tools enables the DF practitioners to perform casework, the role of technology is devoted less attention than the cognitive and human aspects.

As many DF practitioners work at DF units within the police organisation and are, to various degrees, integrated into the investigation teams, the social, structural, and cultural aspects are vital for understanding the success factors and challenges involved when constructing digital evidence. A DF investigation may be described as a sub-process of the overall investigation process, in which the investigation team initiates the work and receives the output. However, due to the scope of the thesis, the social, cultural, and structural dimensions and organisational factors are not explored in depth.

1.2.2 Central terms

The terms *digital forensics*, *digital forensic science*, *digital forensic investigation*, and *digital investigation* are used interchangeably in DF literature and may be assigned different implicit or explicit meanings. For clarity, some central terms for the thesis are explicated.

As a society, we rely on science to make informed decisions about important matters, and forensic science enables us to make similar decisions in courtrooms (M. S. Olivier, 2016a, p. 47). The word *forensic* originates from the Latin word *forensis*, which means pertaining to the forum. The Roman forum was a multidimensional space of negotiation and truth-finding, in which humans, as well as objects, participated in politics, law, and the economy (Weizman, 2014, p. 9). Forensics is used to interrogate the relationship between the *field* – as the site of investigation – and the *forum* – as the place where the investigation results are presented and contested (Weizman, 2014, p. 9). The term gradually came to refer to the court of law, and a

modern understanding of the term *forensics* is “relating to or dealing with the application of scientific knowledge to legal problems” (Merriam-Webster, “forensic”). As forensic science is an applied science, it is argued that it cannot meet the expectations of either an idealised version of pure science or the unrealistic public expectations generated by crime shows and media representation (Julian et al., 2021, p. 92).

DFs is a discipline among the other applied sciences in the forensic science domain. It is a broader concept of what originally was referred to as “computer forensics” and a natural development when more devices than mere computers were included (Pollitt, 1995, p. 1). Performing *digital forensics* or *digital forensic investigation* can be described as:

The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations. (Årnes, 2018, p. 4 referring to Palmer, 2001, p.16)

A *digital forensic science* would thus refer to the development of methods aimed at application in DF investigations and prove their reliability and validity.

A DF investigation should seek to obtain digital evidence in a *forensically sound* manner, which involves preserving the integrity of the evidence if possible. When it is necessary to access evidence in a way that changes some information, the DF practitioner should have the requisite training and experience, and all actions performed on the evidential item should be documented (Casey, 2011c, p. 233). McKemmish (2008, pp. 11-13) defines four evaluation criteria for forensic soundness:

- Meaning: Whether the actions performed during the DF process have changed the meaning of the digital evidence.
- Errors: Whether all errors have been reasonably identified and sufficiently explained as to remove any doubt about the reliability of the digital evidence.
- Transparency: Whether the whole DF process may be independently examined and verified.

- Experience: Whether the DF practitioner handling the evidence has sufficient training and experience.

There have been several attempts to clarify the difference between *digital forensics/digital forensic investigation* and *digital investigation* in the academic literature. In the thesis, the difference relates to the use of scientifically derived and proven methods. It is a digital investigation when the examination and analysis are based on non-scientific investigative methods. An *investigation* is mainly concerned with activities at the pre-trial stage and aims to provide information that may shed light on the six basic questions: what, when, where, why, who, and how (T. Cook, 2016, p. 38).

Digital evidence is a fundamental concept for the thesis and is described in many different ways by organisations and governmental bodies. In the Scientific Working Group on Digital Evidence's (SWGDE) best practice documentation, digital evidence is described as "information of probative value that is stored or transmitted in binary form" (SWGDE, 2016, p. 7). The National Institute of Justice uses the following description: "Digital evidence is information and data of value to an investigation that is stored on, received, or transmitted by an electronic device" (National Institute of Justice, 2008, p. ix). The terms *digital and multimedia evidence* and *digital and multimedia forensics* are used increasingly by organisations such as SWGDE and the National Institute of Standards and Technology (NIST), highlighting the fact that digital evidence not only includes information on computers but also audio files, video recordings, and digital images. The *digital (electronic) device* is vital to the production of digital evidence, and McKemmish (1999, p. 2) describes it as "any electronic device capable of storing information of evidentiary value, including cellular phones, electronic organisers and various network communications devices such as routers and hubs".

Those who have a role in handling the digital evidence during the DF process (or particular stages of the process) are often referred to as DF practitioners in the academic literature. In the thesis, the *DF practitioner* is understood as someone who has DF investigation as the main part of their professional role. When used in the thesis, the term *DF practitioner* does not encompass criminal detectives occasionally performing content analysis of evidence files or patrol officers collecting digital devices as part of incident response and investigation. Yet, it should also be emphasised that being a DF practitioner does not necessarily entail being an

expert in the field. As shown in **Articles 2 and 4**, the background of personnel falling into this category is very diverse in educational type and level, organisational level, and professional experience within criminal investigation and DF work. Some DF practitioners work in specialised units or laboratories within the police, and some may work at independent forensic laboratories. For example, in Norway, each police district has an in-house DF unit, which conducts DF casework in the investigations to which they are assigned. The DF unit also supports and supervises detectives and prosecutors in handling digital evidence in their investigations. Other countries such as the UK and the Netherlands leave parts of the DF casework to external/independent forensic laboratories.

1.2.3 Outline of the thesis

The next section describes the empirical context of the thesis, followed by an outline of the applied theoretical concepts and perspectives that have guided the research. Section 4 describes and discusses the research design, data collection, and applied analysis methods, including a reflection on research quality and ethical issues. Section 5 outlines each article's individual contribution and is followed by a holistic discussion of the results across the papers. Section 6 offers a concluding analysis of the results in relation to the previous research and theoretical assumptions and seeks to sum up the novel empirical and theoretical insights related to the research questions listed in section 1. The thesis's contribution relative to the research question is summarised in section 7, followed by a brief discussion of the implications of the findings for future research, practice, and policy.

2. Empirical context

This section first outlines the brief history of DF and its relationship to the broader forensic science domain and discusses the DF domain's core challenges. The DF process's stages and dynamics are then elaborated in the context of empirical research on DF practice relevant to the particular stage. This is followed by a presentation of research centring on digital evidence. Finally, research on cognitive and human factors in forensic science decision-making is described, followed by a discussion of its relevance for the DF discipline.

2.1 Brief history of DF and its placement in forensic science

The field of DF is relatively new, with a short but complex history. A brief historical outline is essential to understand the DF discipline's standing in today's criminal investigations, as well as the present challenges, including cognitive and human factors. The history of DF did not emerge in a vacuum but relates closely to the implications of technological developments on society, that is, how new technology was adopted and changed people's ways of living, communicating, socialising, and committing crimes. The DF discipline was initially referred to as "forensic computing" (Collier & Spaul, 1992, p. 34) or "computer forensics" (Pollitt, 1995, p. 1). The need for DF developed with the onset of "the digital revolution" (e.g., Floridi, 1999, p. 1) and emerged from the shift from mechanical and analogue electronic technology to digital electronics, which began around the mid-20th century (Computer History Museum, 2021), characterised by the increased production, transmission, and consumption of, and reliance on, information (Holt et al., 2017, pp. 491-526). Pollitt described the history of DF in epochs of pre-history, infancy, childhood, and adolescence (Pollitt, 2010), and the same analogy is adopted here.

2.1.1 Pre-history

During the pre-history (-1985), computers were owned and operated by corporations, universities, research centres, and government agencies (Pollitt, 2010, p. 5), and this world was not significantly networked to the outside world. Technology has been the driving force behind the DF discipline, which evolved as technology continued to influence law enforcement investigations (Holt et al., 2017, p. 430). In the 1970s, individuals discovered methods for gaining unauthorised access to large, time-shared computers, which essentially entailed stealing time on the computers. Many of the new crimes were dealt with using existing laws, but there were some legal struggles, since the law was designed to protect physical property as opposed to intangible digital property (Casey, 2004, p. 25). The early

1980s are considered pre-forensic or ad hoc phases, due to the lack of structure, protocols, training, and tools (Charters, 2009, p. 5). Digital investigations were conducted by law enforcement agents who had received some training in collecting information such as stored data and access logs from computers, and they worked in cooperation with the system administrators to obtain the data. Law enforcement officers were analysing the original evidence rather than a duplicate copy (Holt et al., 2017, p. 435), which reflects not only the lack of technology but also a lack of awareness of the necessity to protect the integrity of digital evidence, during this epoch.

2.1.2 Infancy

During the infancy epoch (1985-1995), an explosion of computer hobbyists emerged - which also included law enforcement personnel - due to the advent of the IBM Personal Computer (PC) (Pollitt, 2010, p. 6). Much of the work centred on recovering data from standalone computers. The internet was not yet well developed, and criminals used dial-up access to compromise computers by so-called “phreakers” (Pollitt, 2010, p. 7; Rogers, 2017, p. 408).

Law enforcement mostly used homemade command-line tools or commercial products, and during this epoch, multipurpose DF tools became available. Some organisations started specialist training programmes (e.g., Electronic Crimes Special Agent Program by the US Secret Service), and the FBI Laboratory established its first Computer Analysis and Response Team (CART) in 1992 (Federal Bureau of Investigation, 1997; Pollitt, 2010, p. 8). While the DF community was growing, very few academics had an interest in this field (Casey, 2019a). Most digital investigations were not yet performed in laboratories or rooted in forensic science principles and methodology (Casey, 2019a, p. 2). Instead, they were typically undertaken by law enforcement officers using personal equipment and with little or no formal training (Pollitt, 2010, p. 8).

Several organisations were established during the infancy epoch, such as the High Technology Crime Investigation Association (HTCIA), the Forensic Association of Computer Technologists (FACT), and the Forensic Computing Group (FCG) (under the auspices of the Association of Chief Police Officers (ACPO)). INTERPOL became a source of information and assistance and provided an international framework for police forces to exchange information, share intelligence, and cooperate at an operational level (Brenner & Schwerha, 2002, p. 359). Collier and Spaul (1992) highlighted the necessity of bringing together

investigative, legal, courtroom, and computing skills to successfully enforce computer misuse legislation. They proposed a new discipline named “forensic computing” (Collier & Spaul, 1992, p. 34), to become an extension of forensic science.

2.1.3 Childhood

During the childhood epoch (1996-2005), there was an explosion of technology for private use, which might be the reason for the period also being referred to as “the golden age” for DF (S. L. Garfinkel, 2010, p. 64). Computers and cellular phones became essential, the internet became “the world’s central nervous system” (Pollitt, 2010, p. 9). Due to the growing volume of technical devices, technical sophistication and legal scrutiny, it became vital to carefully select and train forensic practitioners in the increasingly more specialised field (Pollitt, 2010, p. 10). The field was now driven more by government agencies and professional organisations than individual self-declared professionals (Pollitt, 2010, pp. 10-11). The forensic tools became more sophisticated graphical user interface suites, such as EnCase and Forensic ToolKit (FTK), and the open-source community developed Helix, Sleuth Kit, and Autopsy Browser (Pollitt, 2010, p. 10). In the 1990s, DF started its transformation into a forensic science discipline (Casey, 2019a, p. 2). During this epoch, traditional forensic laboratories began offering digital examinations, and universities worldwide started offering DF courses.

The need for international collaboration and evidence exchange emerged, leading to several initiatives, such as foundation of the International Organization on Computer Evidence (IOCE) (Noblett et al., 2000). In 1998, the High-Tech Subgroup of the G8’s Senior Experts on Transnational and Organized Crime established a 24-7 expert network to offer mutual assistance in high-tech crime investigations, to ensure technical capabilities and legal processes to find criminals abusing technologies and bring them to justice (Brenner & Schwerha, 2002, p. 360; Schjolberg, 2019, pp. 38-40).

An increased focus on the scientific underpinning of the DF discipline initiated a standardisation discourse. The European Network of Forensic Science Institutes (ENFSI) established the Forensic Information Technology Working Group in 1998 (ENFSI, 2022; Geradts, 2011, p. 94). The working group developed best practice guidelines for the examination of digital technology, which have been offered in several updates, with the most recent version from 2015 (ENFSI, 2015a; Geradts, 2011, p. 94). In August 2001, the first

Digital Forensic Research Workshop (DFRWS) was arranged in Utica, New York, which aimed to form a community of interested individuals and start a meaningful dialogue for defining the field and identifying the most important future challenges (Palmer, 2001, p. iii). The first journals dedicated to digital evidence or digital/computer forensics were founded, such as the *Journal of Digital Evidence* (established in 2002) and the *Journal of Digital Investigation* (established in 2004).

By the end of the childhood epoch, computer technology was embedded into every element of daily life. Computer crime and cybercrime, such as online credit card fraud and the distribution of child sexual exploitation and abuse material, was growing rapidly (Pollitt, 2010, p. 9; Rogers, 2017, p. 408).

2.1.4 Adolescence

During the adolescence epoch (2005-), the DF domain has continued to grow in depth and breadth and is faced with more complex challenges to solve. The technology has evolved further, with more rapid internet connections. The availability of social media and smartphones has changed the way people live and socialise into a hybrid online and physical way of living. The IoT and Artificial Intelligence (AI) have made life easier but, at the same time, have introduced new threats and opportunities for criminality. Within the DF domain, the need for expertise and specialised knowledge has increased. In addition to computer forensics, other specialisations and sub-disciplines have emerged, such as live data forensics, mobile device forensics, database forensics, network/cloud forensics, and IoT forensics.

The DF discipline has matured further during this epoch, due to advancements in the scientific underpinning, as well as developments in harmonisation, standardisation, and professionalisation. The American Academy of Forensic Sciences created a new section in 2008 devoted to Digital and Multimedia Sciences (Baker et al., 2013; Casey & Turvey, 2011, p. 259). In the same year, the Forensic Science Regulator was established in the UK to improve forensic science practices in England and Wales, including quality assurance of DF, testing of DF methods, and advancing DF as a scientific discipline (Casey, 2019a, p. 3).

Standardisation has been an essential part of the DF discipline's maturation, and there is now a plethora of available standards from providers such as NIST, the International Organization for Standardization (ISO/IEC), European Telecommunications Standards Institute (ETSI) and ASTM International. Several governmental bodies (e.g., the National Police Chiefs' Council

(NPCC), The Forensic Science Regulator, NIST), inter-governmental organisations (e.g., INTERPOL, Europol, Council of Europe) and non-governmental organisations (e.g., SWGDE, ENFSI, American Academy of Forensic Sciences (AAFS)) are producing and updating normative documents, such as standards, guidelines, and best practices, for the DF discipline. The Cyber-investigation Analysis Standard Expression (CASE) ontology is another example of a recent and important standardisation initiative, aimed to advance the sharing, interoperability, and analysis in cyber investigations (Casey et al., 2017).

Some research efforts have aimed to *harmonise* the DF discipline processes and practices with other forensic science disciplines. A recent harmonisation initiative aimed to identify and describe common core forensic processes from the DF discipline and forensic science domain (Pollitt et al., 2018). Other initiatives have aimed to harmonise DF and forensic science practices for producing evaluative opinions (e.g., Casey, 2020a, 2020b; Horsman, 2021; Tart, 2020).

Researchers and practitioners are also engaged in describing, developing, and codifying the DF process. Since the first process description by the DFRWS workshop, more than 130 process model descriptions have been published in academic journals and conference proceedings, aiming to solve various challenges in DF (Sunde, 2022b). The first process models aimed to provide generic process descriptions for handling DF evidence in a way that made it acceptable to courts (e.g., Carrier & Spafford, 2003; Ciardhuáin, 2004; Reith et al., 2002) and to increase the scientific underpinning of the DF process (N. L. Beebe & Clark, 2005; Carrier, 2006; Casey & Palmer, 2004; Montasari et al., 2015). The continuous process modelling seems to mirror the challenges emerging from the technological developments and new usage patterns. Marsico (2005, p. 2) stated, “These frameworks are more than just guidelines; they also contain the beginnings of what may become theories for the field”. Today, the process models are not only a valuable theoretical underpinning of the DF discipline but also serve as a historical account of the problems that have challenged the discipline over the last two decades.

DF practitioners today are more likely to have academic education, in addition to forensic training, and certifications are often required (Pollitt, 2010, p. 11). There are currently several bachelor’s and master’s programmes aimed at the DF discipline, as well as a few doctoral-

level educational programmes. Nevertheless, there seems to be a general acceptance that the DF discipline is still maturing and has not yet reached the adulthood epoch.

2.2 Challenges of the DF discipline

Although the DF discipline has matured into adolescence, its continuous and rapid evolution has also introduced technical, human-related, organisational, legal, and case-specific challenges. The technical and human-related challenges are elaborated on below, due to their particular relevance to the thesis's research question.

2.2.1 Technical

The first technical challenge is *the volume challenge*, which is caused by more devices among people, more data per device, thus more data per investigation (Casey, 2019a, p. 6; S. Garfinkel, 2012, p. 66; Lillis et al., 2016, p. 11; Luciano et al., 2018, p. 11; Quick & Choo, 2014; Reedy, 2020, p. 29; Ward, 2021, pp. 83-86; Westera et al., 2016, p. 202). According to Moore's Law, the number of transistors on an integrated circuit doubles every 18-24 months. In 1999, a 10-gigabyte hard drive was considered a large amount of data (McKemmish, 1999, p. 5). In perspective, Noblett et al. (2000, 11th para) reported that the readily available systems with 60-gigabyte storage capacity made it practically impossible to examine every file stored on the computer system exhaustively. Today, it is not uncommon to have several terabytes of data in a single case (Bhoedjang et al., 2012, p. 96; Quick & Choo, 2014, p. 277; Wilson-Kovacs et al., 2021, p. 8). Since terabyte is an abstract size description, a more tangible example is found in Haraldseid's master's thesis (2021, p. 37). He interviewed Norwegian criminal detectives about the analysis stage of the DF process. The respondents reported that the analysis could entail the examination of several millions of documents and keyword searches that render hundreds of thousands of hits, requiring weeks of intense work to review. This situation creates performance bottlenecks, and, for example, in the UK, the volume challenge has been met with standardised case prioritisation measures, such as triage (Wilson-Kovacs, 2019) and the National Police Improvement Agency's (NPIA's) High-Tech Crime Unit Case Prioritisation Matrix (Rappert et al., 2021, pp. 6-7). Yet, the volume challenge is not confined to the pre-trial stage. The magnitude of data also complicates case management at the trial stage (Lawless, 2022, p. 200).

The second technical challenge is *the complexity challenge*. It has emerged because of factors such as changing technology, diversity of digital devices, proprietary systems, and encryption

(see, e.g., Muir & Walcott, 2021, pp. 10-11; Raghavan, 2013, p. 92). The complexity challenge causes several problems, such as more time-consuming investigations and lack of standardisation, and introduces a risk of missed opportunities and misinterpreted results (Casey, 2019a, pp. 5-6; Ward, 2021, pp. 108-109). Review of data from serious crime investigations in the South of England established that 50% of enquiries missed all digital investigative opportunities (P. Thompson & Manning, 2021, pp. 112-113). Where a digital opportunity was identified, potential subsequent digital enquiries were missed 47% of the time. A clear indication of the rationale behind the missed opportunities was not identified; however, the authors point towards lack of knowledge or choice as plausible explanations (P. Thompson & Manning, 2021, p. 117). Many DF tools are not equipped to adapt to the rapid changes and evolution of technology, and there is a constant need to develop forensic techniques (S. Garfinkel, 2012, p. 81; Luciano et al., 2018, p. 7). A recent review by NIST suggests that there are hundreds if not thousands of individual techniques that might be used in a DF examination (Lyle et al., 2022, p. 35). The challenge is linked to resources, since, as technology is evolving, there is a continuous need for adequate and up-to-date technology to extract information. As stated by one of the informants in the research of Lawless (2022, p. 196), “I’ve got this problem – I’ve got this crime to investigate but I haven’t got the codec – so I can’t investigate. (Digital forensics practitioner 2014)”. Sometimes, experimentation and improvisation approaches are necessary to extract information from new technology (e.g., Lawless, 2022, p. 200; Ward, 2021, p. 142), and rigorous testing may be required to validate interpretations and assumptions based on examinations of new technology (Horsman, 2019, p. 150).

The volatility challenge is the third technical challenge. Data in live systems and networks may contain information that is valuable to an investigation. At the same time, data may easily be changed, lost or become unavailable during collection, which can harm their evidential value. Yet, there is no way to secure data from a running system without simultaneously changing data (Farmer & Venema, 2005, pp. 5-8, 193-198; Lopez et al., 2016). For example, if the power of a running digital device is turned off during the search and seizure, the content in the random access memory (RAM) may be lost. The volatility challenge is about balancing the obligation to safeguard the authenticity of the evidence against the risk of losing data or data becoming unavailable (encrypted). A study among Norwegian police students showed that many accessed smartphones and other digital devices to search manually for evidence during their year of obligatory practice, often justifying their

actions with the claim that data would otherwise be lost (Andreassen & Andresen, 2020, pp. 76-77).

2.2.2 Human related

The expertise challenge relates to the need for more specialised and advanced knowledge and skills to handle complex evidential sources. At the same time, there is a need for increased capacity to handle the magnitude of digital evidence in criminal investigations. This has led to what is coined as a “decentralization movement”, where personnel with limited knowledge of DF are handling advanced DF tasks in the field (Casey, 2019a, p. 5). Typically, such personnel may not realise the limitations of the methods they use and are incapable of troubleshooting problems with forensic tools, which may lead to missed opportunities and mistakes (Casey, 2019a, p. 6). There are no generally accepted standard minimum competency requirements regarding who is an expert in the DF field throughout the world and no generally accepted curricula for DF education, yet many university colleges and universities provide DF training (Humphries, 2019, p. 40; Vincze, 2016, p. 186). This situation has led to a number of formally qualified experts, all having different levels of competence (Watson & Jones, 2013, p. 826). A systematic literature review of success factors and challenges in DF performed by Cervantes Mori and colleagues showed that lack of formal training, continuous education, and insufficiently qualified staff were frequently mentioned challenges (Cervantes Mori et al., 2021, p. 111). These findings indicate that the challenge of keeping the knowledge up to date at the same pace as the technology developments in society is significant. The expertise challenge also relates to other parties involved in the investigation, such as criminal detectives, prosecutors, and judges in court. Erlandsen (2019, pp. 76-77) found that prosecutors lack the knowledge to adequately challenge the quality of evidence obtained through DF examinations. Technical understanding in courts is highlighted as a challenge by Muir and Walcott (2021, p. 12), since it may lead to requests for further investigation, based on unrealistic timescales or demand for evidence in complicated formats.

The objectivity challenge is a combined human- and law-related challenge, concerned with the normative obligation, within many jurisdictions, to be independent and look for both inculpatory and exculpatory evidence, and is very central to the thesis. In many law enforcement organisations, DF is performed as an in-house service, and DF practitioners are directly involved in the investigation (see, e.g., Andersson, 2020, pp. 24-31; Hansen et al., 2017; Jahren, 2020, pp. 22-23; Lawless, 2022, pp. 195-196). Without formalised

independence between DF and the investigative process, the scientific objectivity of the results may be questioned (Casey, 2019a, p. 7). The question of independence is complicated, because, on one hand, separating forensic capability from the investigation team may ensure independence and prevent exposure to biasing irrelevant contextual information. On the other hand, close collaboration with the investigation team may enable more targeted information gathering and the search for critical information at an early stage of the investigation – informing and potentially eliminating alternative investigative hypotheses and lines of inquiry (Hansen et al., 2017; Jones et al., 2020; Sunde, 2017, pp. 103-104). Whilst cognitive and human factors have been highlighted as a challenge to scientific objectivity and impartiality in other forensic science disciplines (Cooper & Meterko, 2019), the issue has gained little attention in the DF domain. **Article 1** discusses the objectivity challenge and the risk of biased decision-making in DF work. **Article 4** examines whether the DF practitioners used any measures to ensure examiner objectivity during the analysis, and **Article 2** suggests that, despite the DF practitioners’ attempts to be objective, their observations were biased by contextual information.

The quality challenge is a combined human- and organisationally related challenge that might be regarded as a product of all the challenges mentioned above. One of the first to highlight this issue was Casey, in the paper titled “Error, uncertainty, and loss in digital evidence” (2002). Influential reports such as the National Academy of Sciences (NAS) report (2009) highlighted the lack of scientific rigour and the risk of confirmation bias in forensic science, including the DF discipline (National Research Council, 2009). Cervantes Mori et al. (2021, p. 111) uncovered several issues related to the quality challenge, such as lack of standardisation, insufficient quality management, missing tool validations, irreproducibility of examinations, lack of scientific validation, and inconsistencies in terminology. The continuous strive towards catching up with the technology speed train may have caused a sharp efficiency focus, leaving quality issues in the background. The professional cultures and structures from whence the DF discipline originated may also be a confounding variable to the quality challenge. Atkinson (2014, pp. 253-254) comments that the software engineering culture is founded on a business model welcoming secret source code and patents, aimed at profit rather than open peer review and scientific validity, which may explain why DF differs from other forensic science disciplines concerning the scientific rigour in methods and practices. **Articles 4** and **5** shed light on the quality challenge, by exploring whether the DF practitioners applied techniques to control evidence reliability during the analysis and whether

they provided sufficient and accurate documentation concerning the applied tools, methods, and procedures in their reports.

When considering the technical and human related challenges in context, the police's capability and capacity to investigate crimes involving digital evidence is also a *legitimacy challenge*. The police must be able to secure evidence effectively to prevent, detect and clear crime. At the same time, they must act in compliance with legal requirements, quality standards, and ethical frameworks and ensure minimal intrusion into private data (National Police Chiefs' Council, 2020, p. 6).

2.3 DF practitioner conduct

The thesis's research question centres on DF practice and the role of the DF practitioner in the construction of digital evidence. This section summarises the empirical research on how DF work is performed.

The continuous technology development and the associated technical challenges seem to have been a significant driver of the research within the DF domain. A large body of research has thus been concerned with developing new methods, tools, processes, procedures, and frameworks for handling new technology or novel implementations of technology. In contrast, relatively few empirical studies have examined DF practice. These few studies that exist are based on empirical data, mainly collected through methods such as observation, interviews and surveys. As will be shown, providing a generalised, all-encompassing, and valid description of how DF work is enacted is challenging (if not impossible), due to the diverse nature of how DF work is organised and performed, the lack of a formalised and binding standard, and the sparse body of empirical research on DF practice. The DF process model described by Anders Flaglien (2018) was used in the analysis in **Article 1** and is also applied here as a framework for describing the typical tasks performed during the stages of the DF process. The DF process is a simplified, generalised, and idealised outline of the stages of handling digital information in a criminal investigation. The relevant empirical research on DF practice, from both social science researchers' and DF scholars' perspectives, is discussed at each stage. This will shed light on the research gaps concerning cognitive and human factors in DF casework, particularly at the analysis and presentation stages, where the thesis offers novel insights. The thesis's empirical findings are not integrated here, and but will be presented and discussed in sections 5 and 6.

2.3.1 Identification, collection, and examination stages

During the *identification* stage, the DF practitioner aims to identify digital devices or systems that might contain information relevant to the case either present at the search scene or at other physical or virtual locations. When identified as relevant, the evidence must be preserved. This is done by isolating, securing, and documenting the physical and digital evidence (Flaglien, 2018, pp. 18-19). During the *collection* stage, the data from digital devices or spaces are acquired, which means copying, if possible, bit-by-bit, using appropriate methods and techniques (Flaglien, 2018, pp. 25, 149). When the digital devices are brought back to the DF laboratory, the DF practitioner acquires the data from the digital device. The acquisition process and procedures depend on the device type, acquisition scope, and which data are considered a high priority. The procedure follows generally accepted forensic principles to preserve evidence integrity if possible. In instances where it is not possible to collect data without compromising the integrity, for example when acquiring data from a live system, the collection should be performed in a manner that minimises alteration and does not change the meaning of the secured information (McKemmish, 2008, pp. 10-11). The order of volatility must be taken into account, ensuring that most volatile data are acquired before less volatile data (Flaglien, 2018, p. 31). The collection should be performed by DF practitioners with sufficient expertise, who should be transparent about the actions performed on the device (Farmer & Venema, 2005, pp. 5, 193; McKemmish, 2008, pp. 12-13).

During the *examination* stage, the raw data are prepared for subsequent analysis through restructuring, parsing and pre-processing (Flaglien, 2018, p. 33). The DF practitioner performs various forensic procedures and tasks, for example opening containers with compressed files (e.g., zip files), decrypting encrypted file containers, recovering deleted files, and verifying file signatures. DF software automates many of these tasks, but manual examination is sometimes necessary, due to software limitations and the necessity to control for software interpretation errors.

A few practice-oriented empirical studies have focused on the identification, collection, and examination stages. These studies have examined procedures and processes such as acquisition tasks (Carlton, 2007; Hewling, 2013) and preview (J. I. James & Gladyshev, 2013b). Wilson-Kovacs and colleagues conducted an ethnographic study, which explored the prioritisation and triage practices of DF practitioners in four English constabularies (Wilson-Kovacs, 2019). Andersson (2020) explored the practices concerning the chain of custody of mobile phone exhibits in two Swedish police districts.

Research has shown that acquisition initially was a task performed solely by DF practitioners. Today, patrol and investigation officers often conduct this task, due to the implementation of automated solutions or kiosks facilitating the acquisition of data from, for example, smartphones (Andreassen & Andresen, 2020; Collie, 2018; Wilson-Kovacs, 2019). Ward (2021, p. 106) found that, due to the rapidly changing technology, the acquisition methods often did not work adequately on specific data types. The practitioners needed to improvise to secure data, which led to more lengthy investigations.

The research conducted by Wilson-Kovacs and colleagues (2019) relates largely to the examination stage. It sheds light on how the demand for digital information, due to its availability and usefulness as evidence, combined with the volume challenge for the DF domain, has led to the need for measures such as triage to prioritise and eliminate irrelevant exhibits. Still, triage does not solve the issue alone, since triage tools may have limitations and fail to identify relevant information due to not being up to date on the latest technology (Rappert et al., 2021, p. 6). A small experimental study by J. I. James and Gladyshev (2013b) sheds light on another challenge with triage. Five DF practitioners previewed evidence files and were asked to decide which to include or exclude for further investigation. The results showed that the participants were inconsistent in their decisions on which exhibits to include or exclude. This research shows that the triage tools support human decision-making and that these decisions may be inconsistent. The consistency in practitioner decision-making was further examined in the thesis's **Article 2**.

2.3.2 Analysis stage

At the *analysis* stage, an in-depth analysis of the information is performed in light of the assignment or mandate given to the DF practitioner. Typically, the analysis entails reviewing large volumes of information, and the DF practitioner uses different approaches to filter out irrelevant information, for example by focusing on a defined time period or type of information, such as images or documents, or targeting relevant information, for example through keyword searches.

Investigative strategies and analysis categories

Different strategies may be applied for *investigative reconstruction*, which Casey and Turvey (2011, p. 255) describe as “the systematic process of piecing together evidence and information gathered during an investigation to gain a better understanding of what transpired between the victim and the offender during a crime”.

The analysis can be divided into three categories. The first is *technical analysis*, focusing on the reconstruction of events based on traces on the evidence file. Three fundamental types of reconstruction are outlined in the DF literature: *functional analysis* centres on the functions of the computer system; *temporal analysis* examines the time and sequence of events and, finally, *relational analysis* investigates relations between entities, such as people, email addresses, aliases, IP-addresses, and telephone numbers (Casey, 2011a, pp. 499-506; King, 2006, pp. 22-23). *Content analysis* is the second analysis category and entails assessing and selecting content relevant to the case under investigation, such as images, chat conversations and documents (Sunde, 2017, p. 25). The third analysis category is *evidence evaluation*, which involves determining the evidence's value or strength.

The distinction between the *investigative* and *evaluative* approach is important at the analysis stage. The evaluative approach is concerned with establishing the value or strength of the uncovered findings in light of a set of propositions and conditioning information (Casey, 2020b; ENFSI, 2015a, pp. 36-39; 2015b; Pollitt et al., 2018, p. 9). Evaluation should follow a structured procedure (ENFSI, 2015a, pp. 34, 41; 2015b), and it is argued that evidence evaluation requires a higher degree of expertise than providing investigative opinions (Casey, 2016, p. A2). There are several research papers describing and promoting formal evaluations of digital evidence (Casey, 2020b; ENFSI, 2015a, 2015b; Horsman, 2021; R. Overill & Chow, 2018; R. E. Overill & Collie, 2021; Pollitt et al., 2018; Ryser et al., 2020; Sunde & Horsman, 2021; Tart, 2020). Yet, there is little to no research on how such evaluations are performed and documented in DF practice.

In contrast to the available descriptive and normative guidance for the analysis stage, the empirical research relevant to investigative strategies and analysis categories is sparse. Ward (2021) interviewed DF practitioners about their practices when extracting and analysing data from mobile devices. Haraldseid (2021) examined practices for the content analysis of evidence files in two Norwegian police units and described content analysis as a selection of relevant information that covers the purpose of the investigation, including a foundation to decide the question of the indictment (p. 19). The study suggests that content analysis is also a speciality that requires a combination of skills, such as expertise in using the analysis software: bookmarking, filtering, creating adequate search phrases, and understanding its limitations. Haraldseid's study found no standard or defined analysis method used for content analysis among the participants.

Brookman and colleagues conducted a four-year ethnographic study of the use of forensic technoscience in 44 British homicide investigations (the offences took place between 2011 and 2017), which included mobile phones, computers, and CCTV evidence. They produced a series of papers based on the study (Brookman & Jones, 2021; Brookman et al., 2020a, 2020b; Innes et al., 2021; Jones et al., 2020), which are relevant to the thesis, since they partly relate to the role of digital evidence and the DF practitioners. Their research describes how actors socially construct the meaning and significance of data derived from digital devices through narrative development.

A recent study by NIST, named the “Black-box study for digital forensic examiners” (further referred to as “the NIST black-box study”), aimed at measuring the performance of DF practitioners from both the public and private sectors (Guttman et al., 2022). The authors conclude that, despite the study’s limitations, “it demonstrated that digital forensic examiners could answer difficult questions related to the analysis of mobile phones and personal computers” (Guttman et al., 2022, p. 1). However, the results reveal that the proportion providing incorrect answers to questions rated as *basic* for the mobile device image ranged from 0% to 51.9%, and from 0% to 34.3% for the computer hard disk image.

This shows that although some studies have explored DF practitioners’ performance at the analysis stage, there is a knowledge gap concerning how investigative strategies, analysis categories, and evidence evaluations are performed in practice. The thesis aims to expand the empirical insights concerning these issues in **Articles 2-5**.

The role of hypotheses at the analysis stage

The analysis stage’s objective is to find information that may support or refute investigative hypotheses or form the basis for new hypotheses (Flaglien, 2018, p. 40). At an early stage of an investigation, the digital information may help to generate hypotheses about what has happened and whether the incident is a crime or not. Digital evidence may also support or refute the existing hypotheses or verify the validity of other information (Ekfeldt, 2016, pp. 269-271; Flaglien, 2018, pp. 17-18).

The role of hypotheses is a recurring theme in theoretical DF research papers, but two different perspectives seem to be involved. One perspective is related to advancing the scientific underpinning of DF, and Carrier (2006) was one of the first to highlight the role of

hypothesis forming and testing for this purpose. However, it is argued that the hypothesis-based approach is one of the methodological weaknesses in the current meta-theory of DF (Tewelde et al., 2015, p. 30). Tewelde et al. underline that not every hypothesis is *per se* scientific, and the demonstration of a hypothesis-based approach does not directly guarantee the scientificness of a discipline (Tewelde et al., 2015, p. 30). They refer to Windelband's (1998) terms and emphasise that the hypotheses formulated in DF are *ideographically* (towards the unique) rather than *nomothetically* oriented (what is universal/general) – where the latter characterises scientific hypotheses. They warn against an illusion of scientificness in DF work that may arise due to a lack of awareness of the many different classes of empirical hypotheses (Tewelde et al., 2015, pp. 37-38).

The other perspective involving hypotheses relates to the view of DF as an investigative tool and a sub-process for the overarching criminal investigative process. Here, the evidence files are reviewed for traces and content that may provide investigative leads or may be used as evidence in court. The role of such hypotheses seems to relate to the degree to which the hypothesis-driven investigation is implemented as an investigative strategy, such as in Norway (Politidirektoratet, 2017; Skre, 2020).

The empirical research concerning the use of hypotheses in DF casework is limited to a few studies. The issue was, to some extent, investigated in my master's thesis (Sunde, 2017, p. 79), which showed that the interviewed DF practitioners did not themselves generate hypotheses as a basis for their analysis but were sometimes guided by the investigative hypotheses defined by the investigation team. The criminal detectives interviewed by Haraldseid (2021, pp. 53-56) referred to investigation plans and hypotheses as a tool for underpinning objectivity in content analysis and safeguarding the suspect's rights. However, their descriptions of practice showed that the official guidelines were not always operationalised into a hypothesis-driven content analysis. **Articles 4 and 5** provide insights into how hypotheses are developed, used, and documented during DF investigations.

Collaboration and information exchange

Close collaboration during the DF process between general and technical investigative competencies is highlighted as a success factor in several research articles (e.g., Cervantes Mori et al., 2021, p. 108; Hansen et al., 2017, p. 9; Leppänen & Kankaanranta, 2017, p. 168). Empirical research on collaboration practice is limited to a few studies which explored the collaboration between criminal detectives and DF practitioners in the Netherlands and Norway (Borhaug, 2019; Hansen et al., 2017; Sunde, 2017) or tensions and professional

dynamics between Digital Media Investigators, DF practitioners, and criminal detectives in England and Wales (Rappert et al., 2021).

The thesis sheds light on some of the disadvantages of close collaboration. **Article 1** discusses the possible biases that may affect DF work. **Article 2** shows that contextual information is often forwarded to the DF practitioners when assigned to the case through submission forms or dialogue around the assignment. **Article 4** shows how this information may influence the DF practitioners' analysis approach, in terms of hypotheses generation prior to starting the analysis, and **Article 2** shows that the contextual information may bias the observations during the analysis of digital information.

2.3.3 Presentation stage

During the *presentation* stage, the DF process and the results are documented, often in multiple reports. The technical procedures, methods, and tools used for handling, collecting, acquiring, and processing the data are also reported for transparency and auditability in DF casework. A detailed record of who has handled the evidence from the time of collection is created to prove the chain of custody and is maintained until the case is closed (Flaglien, 2018, pp. 46-47). The relevant findings are described and documented in an analysis report through technical, investigative, or evaluative reporting styles (ENFSI, 2015a, pp. 40-41; Horsman, 2021). The reports and the results may undergo quality control before the report is submitted to the client. At the trial stage, depending on the jurisdiction, the DF practitioner's reports are presented as evidence in court, or the DF practitioner is called to the court proceedings to present and explain the findings from their reports orally. Documentation is vital for describing the evidence, explaining its relevance, and evaluating its value to the case under investigation. Documentation is also essential for demonstrating a forensically sound DF process, preservation of evidence integrity, and an unbroken chain of custody.

The research related to the presentation stage is mainly theoretical research or technological developments for a more efficient reporting stage. Several research papers are concerned with tools or systems for automated reporting and the need for standardisation in reporting (e.g., Karie et al., 2019; van Beek et al., 2020). Although several guidelines provide recommendations about what a report should contain (e.g., ENFSI, 2015a, pp. 40-41, 61; INTERPOL, 2019, p. 53), the formal reporting guidance targeted at the DF discipline is relatively sparse (Horsman, 2020, p. 627).

Efficiency is a recurring theme in academic discussions about reporting. Streamlined forensic reporting (SFR) was introduced in England and Wales to achieve “swift and sure justice” through an early agreement with the defence on forensic issues or to identify contested issues where this cannot be achieved in the first instance (McCartney, 2019, p. 83). Nevertheless, the SFR’s ability to facilitate sure justice is debated, due to the de-skilling of those involved and the risk of mistakes and misinterpretations of the results. “SFR may be preventing defendants from mounting a proper defence, but also defendants are being charged/convicted on the basis of flawed, or at least, incomplete scientific evidence” (McCartney, 2019, p. 85).

A study comparing the quality regimes in the DF, DNA, and fingerprint disciplines in the UK found that the DF discipline “is operating under arguably less rigorously defined standards, practitioner governance and evidence validation procedures” (Page et al., 2018, p. 84). To ensure sufficient quality and to mitigate errors, it is suggested that the report (and results) should undergo quality control and peer review at an appropriate level and scope (Horsman & Sunde, 2020; Page et al., 2018, pp. 90-92; Sunde & Horsman, 2021). However, standardisation has been debated within the DF community. Some have promoted it to safeguard the credibility of the field and ensure the admissibility of evidence in court (e.g., Grobler, 2012; Guo & Hou, 2018; Marshall & Paige, 2018; Meyers & Rogers, 2004; Zahadat, 2019). Others have been sceptical about standardisation as the best solution to challenges, due to the cost and effort involved in complying with the standards and the complexity of validation in an environment of rapidly changing technologies (see, e.g., Sommer, 2010).

The empirical research relevant to this stage has mainly been concerned with the quality control of DF casework. Tully et al. (2020) investigated the implementation of and compliance with quality standards in England and Wales. The current requirement is for organisations carrying out DF to gain accreditation to the international standard ISO/IEC 17025 and the Forensic Science Regulator's Codes of Practice and Conduct, which fosters a systematic approach to quality. Tully et al. (2020) reviewed the available data from initial assessments and surveillance visits to accredited units by the United Kingdom Accreditation Service (UKAS) from 2015-2019 and quality referrals to the Forensic Science Regulator between 2012 and 2019. The aim was to determine whether the ISO/IEC 17025 standard addressed issues in DF and identify factors that could assist the implementation of quality systems. Tully et al. concluded that the study supports the need for quality standards in DF and that accreditation to standards gives external assurance that an organisation has the

sustainable competence to produce reliable results in the accredited activity. At the same time, they acknowledge that standards are no guarantee for quality and the elimination of all errors, partly since the quality standard does not address the financial viability of a company (Tully et al., 2020, pp. 9-10).

Jahren (2020) explored the state of quality assurance in three Norwegian police districts. Similarly to Tully et al.'s (2020) results from the initial assessment, Jahren found that peer review of reports was rarely performed. When performed, it happened in an informal manner, where a colleague asked another to look at the report. Correspondingly, the survey conducted by Haraldseid (2021, pp. 45, 61) showed that the result of the content analysis would not routinely undergo peer review.

In a survey among Dutch police using the investigation platform Hansken, Borhaug (2019, p. 47) found that the system provided technical validation but that the conclusions had to be validated by the DF practitioners. Of DF practitioners, 70% replied that they had verified investigators' work, but only 11.8% said they did this routinely (every time). Unfortunately, the survey did not examine whether DF practitioners conducted peer reviews of each other's work.

Stoykova and colleagues examined DF reports obtained from 21 Norwegian criminal cases and found substantial deficiencies in documentation practices related to the reliability and chain of custody of the digital evidence (Stoykova et al., 2022). They concluded that, in most of the examined cases, it was impossible to trace the DF actions performed or to link the digital evidence to its source based on the available documentation. The reports collected through the DF experiment expand these insights and shed light on how the assignments and procedures, including any quality measures, are documented (**Articles 3 and 5**) and how the results are presented (**Articles 3 and 5**). In contrast to Stoykova et al.'s (2022) study, the DF experiment was not limited to Norwegian DF practitioners.

2.3.4 Process dynamics and dependencies

Although the DF process is outlined here as linear, it is in fact *iterative* due to the necessity of revisiting former stages (Flaglien, 2018, p. 16; Horsman & Sunde, 2022, p. 177). For example, traces of connected devices (e.g., an external hard drive) may be discovered during the analysis, which might require further identification and collection. Also, the peer review

can uncover uncertain, incomplete, or poorly justified findings, which might require a revisit to the examination or analysis stages for further investigation or verification of the findings.

The DF process may be characterised as a *multiple device process*, since a person would typically have several digital devices, and some may be seamlessly interconnected. For example, a user may have a smartwatch connected to her smartphone, which is connected to cloud storage. The smartwatch and smartphone are connected to the internet and telecom networks, leaving traces when the devices are in use or simply through being carried around while turned on. The user does not usually control the transfer of information between these devices, where it is stored, and which external network devices they communicate with – and might not even be aware of the process. Hence, as the relevant traces to an investigation are distributed to multiple devices, they must be collected to get a complete overview of, for example, the suspect's activities or movements (Sharma et al., 2020).

Due to the necessity for *various types of expertise*, many people with different roles, competencies, and epistemic cultures are involved in transforming the information into meaningful evidence in a legal context (see, e.g., Collie & Overill, 2020). The research by Brookman and colleagues highlights the collective and social nature of sense-making in British homicide investigations (Brookman & Jones, 2021; Brookman et al., 2020a, 2020b; Innes et al., 2021; Jones et al., 2020). They describe how actors socially construct the meaning and significance of digital evidence, such as CCTV footage, mobile telephone cell site data, and other evidence types, by sequencing and arranging characters and events into a temporally ordered storyline of a crime (Brookman et al., 2020a, p. 22).

A few studies have shed light on some of the disadvantages or risks of involving several people and different competencies in the investigation. A US-based study of DF practitioners' practices with investigations involving mobile devices showed that engaging multiple investigators also entailed more lengthy investigations (Ward, 2021, p. 82). The study by Stoykova et al. (2022, pp. 10-11) showed that investigative and forensic activities were often performed in parallel, resulting in difficulties in tracing the DF actions performed on each item and linking the digital evidence to its source. Digital devices are not necessarily collected by DF practitioners but, instead, by patrol officers (Andreassen & Andresen, 2020; Harrison, 2004, pp. 81-82). As shown by Wilson-Kovacs (2019), triage is often performed by investigation officers, and is aimed at eliminating irrelevant exhibits. Senior investigating officers would act as gatekeepers, deciding which exhibits to forward to the DF unit for in-

depth analysis. While acquisition used to be a task for DF practitioners, self-service kiosks for the automated collection of content from smartphones and other handheld devices are becoming more widespread in law enforcement organisations (Lawless, 2022, p. 198; National Police Chiefs' Council, 2020; Rappert et al., 2021, pp. 6-10). After collection and examination, the DF practitioner may perform the analysis on their own, collaborate with the criminal detective on the analysis, or facilitate the content analysis, which is undertaken by the criminal detective (Sunde, 2017, p. 64). In-depth knowledge of the particular case and general knowledge of the investigated crime phenomenon are highlighted by DF practitioners as a prerequisite for deciding what is relevant information or not during the analysis (Sunde, 2017, p. 62). A criminal detective would thus often be involved in reviewing the information, and the DF practitioner would, in such situations, prepare and facilitate the review (Hansen et al., 2017; Horsman & Sunde, 2022, p. 176; Sunde, 2017, p. 64). The information deemed as relevant would sometimes need a deeper technical analysis, which would require the DF practitioner's expertise.

There are typically *multiple hardware and software tools* involved in the DF process specialised for the device, technology, type of data, or the scope or nature of the task to be performed to uncover relevant evidence (see, e.g., Narwal & Goel, 2020). The tools aim to facilitate effective DF investigations and provide credible results. Specialised technology and software are necessary for triaging devices for potentially relevant information, acquisition of the digital content, to pre-process the information for the type of content targeted by the investigation (such as images and documents), and to facilitate efficient review.

Tools need to be updated frequently to handle new technology and are thus vulnerable to programming flaws, which may lead to systematic errors in the output (Horsman & Sunde, 2020, pp. 2-3; SWGDE, 2018, pp. 3-4). A comparative study of mobile forensic toolkits showed considerable variation between recovery methods implemented in the toolkits, the proportion of recovered artefacts and the extent to which the results produced by one recovery method can be verified by one or more others (Glisson et al., 2013, p. 55). The dual-tool approach is a quality measure for uncovering tool interpretation errors and involves checking the interpretation with a second tool (Flaglien, 2018, p. 32). However, the dual-tool approach has several limitations (Friheim, 2016). Multiple overlapping tools are also necessary to compensate for the shortcomings in other tools (Ward, 2021, p. 86). An aspect of tools that is essential to human factors is that the tools influence how the investigation is performed, due,

for example, to how well the tool fits the purpose (J. I. James & Gladyshev, 2013b, p. 156). Factors such as functionalities, interface, performance, and how the results are presented and ordered may influence or bias the investigation (Dror & Mnookin, 2010; Haraldseid, 2021, pp. 45-47; Lawless, 2022, p. 196).

The volume and expertise challenges have led to the decentralisation movement and the increase of so-called *push-button forensics*, with highly automated tools for the DF process. On one hand, this approach increases DF capacity, since it allows less technically skilled personnel to conduct tasks (J. I. James & Gladyshev, 2013a, p. 14). On the other hand, there is a risk of misinterpretation of the results, due to inadequate technical competence (Collie, 2018; Humphries et al., 2021, p. 9).

2.4 Digital evidence agency - increasing volumes and new functions

As mentioned in section 1.2, Innes et al. (2021) define three distinct framings of criminal investigation research: crime, conduct, and techniques. However, drawing on perspectives from actor-network theory (ANT), where both human and non-human are perceived as actors with agency in the network of information processing and knowledge production (Latour, 2005), one additional framing could be added – the social life and role of the digital evidence. The social life of digital traces, and specifically how human and non-human actors influence the construction of the digital evidence, is given particular attention in **Articles 3 and 5**. Scientific practice and knowledge construction have been the subject of several empirical studies from STS scholars (e.g., Knorr-Cetina, 1981; Latour & Woolgar, 1979/1986), and some studies have centred on the construction of forensic evidence (e.g., Cole, 2001; Costa & Santos, 2019; Dahl, 2008, 2009; Kruse, 2016; Williams & Weetman, 2013). Yet, these studies did not encompass digital evidence.

While examining investigative practices, the social life of digital evidence has been covered to some extent by Brodeur (2010), who examined 153 Canadian homicide case files from the period 1990–2001. He found that digital evidence (electronic or physical surveillance) and computer searches was a determinant for identifying or locating suspects in a very small proportion (0.7-2.6%) of the cases (Brodeur, 2010, pp. 208-210).

Although the framing of Brookman and colleagues' ethnographic research was the investigative conduct in 44 UK homicide cases, digital evidence is discussed in all published papers from the project to date (Brookman et al., 2020a, 2020b; Innes et al., 2021; Jones et al.,

2020). One paper devotes particular attention to CCTV evidence, and in contrast to Brodeur (2010), Brookman and Jones (2021, pp. 8-9) found that digital evidence in the form of CCTV and phone data was the most important factor for both identifying and charging suspects. Evidence, such as social media and computer examinations, played a role in only a few (1-2) of the examined cases. Nevertheless, evidence types such as mobile phone extractions and browser history logs are highlighted as critical by several of the interviewed homicide detectives for shedding light on the motive and intentions (Innes et al., 2021, pp. 714, 717). The increased importance of digital evidence is no surprise, given how society has been digitised during the period between these studies.

2.5 Cognitive and human factors influencing forensic decision-making

There is now a substantial amount of research on cognitive factors influencing forensic work. A key aspect of this research has been to explore whether and how irrelevant contextual information may influence perception, judgement, and decision-making within the forensic casework. Since three systematic reviews of key relevance to the thesis were recently published, it was considered unnecessary and redundant to perform an additional review for the thesis. The first was a systematic review of cognitive bias research in forensic science, performed by Cooper and Meterko (2019). The review encompassed 27 studies from the following disciplines: fingerprint, forensic anthropology, bite-mark, bloodstain, dog handling, DNA, hair, handwriting, shoeprint, speech, tool marks/bullets, crime scene investigation, and forensic pathology. Two studies about human-technology interaction related to the Automated Fingerprint Identification System (AFIS) fingerprint system and face matching were also included. Cooper and Meterko concluded that the studies underpin the fact that case-specific information, even when it is wrong, can influence forensic decision-making. They highlight the use of ancillary information (information of other types of evidence) in the decision-making as a biasing source for forensic decisions and strength of conclusions. They also underline the need for research on the algorithm-generated matching process of human decision-making (Cooper & Meterko, 2019, p. 43).

The second literature review was conducted by Kukucka and Dror (2022) and included 43 empirical studies on cognitive bias in the forensic science disciplines and domains. In addition to the disciplines included in the above-mentioned study by Cooper and Meterko (2019), they also included the studies from disciplines concerned with arson, toxicology, and DF (**Article 2**). The review showed that although 6 of the 43 included studies did not find statistically

significant cognitive bias effects, such effects were found across all the included forensic disciplines.

The third systematic review, which was also performed by Meterko and Cooper (2021), included 30 empirical social science research papers on cognitive biases in criminal case evaluations. Since the DF discipline may be considered both a forensic science discipline and an investigative tool (Casey, 2013, p. 86), research on cognitive biases in criminal investigations is also relevant. Meterko and Cooper summarise that the body of research demonstrates that the human element can unintentionally undermine the truth-seeking in the evidence integration process (Meterko & Cooper, 2021, p. 10).

The literature review for the thesis shows that, although decision-making is a recurring theme in DF scholarly literature and research, little of this stems from empirical research on decision-making. Instead, they are mostly presentations of opinions or ideas about how to make effective decisions in DF work (e.g., Bednar et al., 2008; Grigaliunas et al., 2021; Horsman, 2019), how technology such as AI, machine learning, or data mining may assist in such decisions (e.g., Böhm et al., 2021; Costantini et al., 2019), or how technology may perform the DF decision-making through an automated process (e.g., Kelly et al., 2020). However, three smaller experimental studies and one qualitative study with relevance to DF decision-making (not included in the above mentioned reviews) were identified. In the first study, J. I. James and Gladyshev (2013b) engaged five DF practitioners in an enhanced preview of evidence files. The results showed that the participants were inconsistent in their decisions about what to include or exclude based on the preview. In the second study, nine DF practitioners participated in a comparative study of visualisation processes for analysis, which aimed to enhance their decision-making and facilitate the explanation of phenomena in evidentiary data (Osborne et al., 2012). They found that the EPIC (Explore, Investigate and Correlate) process visualisation approach led to the best performance and user satisfaction, compared to two other visualisation approaches. Although not involving DF practitioners, the third study examined the reliability of decision-making among five analysts when classifying child sexual exploitation material (Kloess et al., 2021). The study showed that the level of agreement on age estimation was moderate to good, and very good for image classification. The fourth study explored forensic analysts' self-reported decision strategies when identifying and investigating cyber intrusions. Based on interviews with nine forensic analysts specialised in triage, incident response, or forensic examinations, Sanders (2021) applied cognitive task analysis methods and developed a model of diagnostic inquiry that represents the

relationships between how analysts formed investigative questions, interpreted evidence, assessed the disposition of events, and chose their next investigative actions.

To summarise, there is a knowledge gap concerning DF decision-making. **Article 1** contributes to bridging the research about the influence of cognitive and human factors on decision-making from forensic science disciplines into the DF domain. **Article 2** is the first study of bias and reliability in DF observations, interpretations, and conclusions. It indicates that contextual information may bias DF observations, and **Article 4** sheds light on how the contextual information may influence the DF practitioner's beliefs and hypothesis generation prior to analysing the evidence file.

3. Theoretical perspectives on the DF practitioner's role in the journey from digital trace to evidence

Theory is a way of thinking systematically about concepts or what things are, mechanisms or structures related to how things work, or normative assumptions about how things should be (Nygaard & Solli, 2020, p. 131). This section outlines the main theoretical concepts and perspectives relevant to the five articles in the thesis and the concluding analysis and discussion in section 6. It centres on theoretical concepts relevant to the analysis and presentation stages of the DF process. The thesis builds on the foundation of the empirical research summarised in section 2.3, however, new perspectives are also applied. Outlined first is the concept of the trace, which is the starting point of the evidence construction process. The digital trace is essential to the analysis and discussion in **Articles 2-5** and therefore explained in detail. Then, the focal point moves to the human factor, and theoretical perspectives of how DF practitioners turn traces into evidence are outlined. The scientific inquiry and investigation perspectives particularly relate to analysis and discussions in **Articles 1, 2, 4 and 5**, whilst perspectives concerning inscription practices and narrative construction are used as analytical concepts in **Article 3**. Then, the normative assumptions of what makes up evidential value in a legal context are outlined, which are central for the analyses and discussions in **Articles 2, 3 and 4**. Finally, the concept of error, which concerns all articles, is described. Particular attention is directed towards cognitive factors such as bias, biasing sources and reliability, which are the primary focus in **Articles 1 and 2**.

3.1 Traceology

The thesis explores the DF practitioner's role in constructing digital evidence within a criminal investigation. A criminal investigation is concerned with reconstructing past (or still ongoing) events and shedding light on all necessary aspects relevant to a criminal trial. This entails obtaining sufficient information to construct a plausible narrative about the event and the persons involved. Since the police have no direct access to what actually happened, they will have to rely on traces resulting from these events and use different types of logic and investigative processes, methods, and tools, to be able to infer with sufficient confidence what the traces are, what activities or events caused them, and who was involved. **Article 2** provides novel insights into issues concerning observing the trace in the first place, considering the traces in context, and drawing inferences and conclusions based on these observations. The trace does not represent itself in court, and **Articles 3 and 5** shed light on

another dimension – namely, how the written representation of the trace in reports may introduce “evidence elasticity” in what the trace is, what it means, and its evidential value to the case under investigation. To understand how digital evidence is constructed and the components that make up digital evidence, one must take a step back and explore a more fundamental concept – *the trace*. Since the trace may be understood from different perspectives, it is first described from a perspective inspired by Charles S. Peirce’s semiotics, followed by a perspective inspired by the STS tradition.

3.1.1 Theorising the trace from a semiotic perspective

Different meanings have been used when describing the trace. According to Tilstone et al. (2013, p. 177), two meanings have frequently been assigned: The first is related to the *semiotic* tradition, which mainly was developed by Peirce and describes the trace as a mark, object, or other indication of the existence or passing of something. The second is more focused on the *quantity* of the object – and describes a trace as a tiny quantity, too small to be accurately measured. For example, the chapter “Trace evidence” in the *Handbook of Forensic Science* relates trace to the latter: “Trace evidence is a category of evidence that is characterised by the analysis of materials that, because of their size or texture, are easily transferred from one location to another” (Houck, 2009, p. 166). Whilst quantity is an issue in the forensic science disciplines dealing with physical traces, it is not helpful in the digital world, since the quantity does not necessarily reveal anything significant about the event. The thesis thus draws on the semiotic understanding of the trace.

Pollitt et al. (2018, p. 1) underline the relationship between a trace and an event. They state: “A trace is any modification, subsequently observable, resulting from an event”. They argue that all traces involve some modification, which affects either an entity in an environment or the environment itself. This understanding entails that immutable objects can also be considered a trace when their occurrence is the consequence of an event, such as a mobile device identifier deposited at a crime scene. The trace can be a presence or an absence. Its nature can be physical or virtual, material or immaterial, analogue or digital (Pollitt et al., 2018, p. 1).

Whilst also drawing on the semiotic view of the trace, Jaquet-Chiffelle and Casey (2021, p. 2) aim to develop a formalised model of the trace represented by mathematical terms, which can cultivate a unified understanding of the trace across forensic disciplines. They take the pragmatic stance and relate the work to Peirce’s concept of sign chains with dynamic and

immediate objects. Here, they distinguish between the tangible and the abstract world, whereas, in forensic science, the abstract world is the hypothetical former state in which the former events occurred or not (Jaquet-Chiffelle & Casey, 2021, p. 6). In their model, what makes up a trace is not the observable object itself, as referred to by Pollitt et al. (2018, p. 1), but what it represents (perceived differences) relative to inferences about the abstract world (Jaquet-Chiffelle & Casey, 2021, p. 9). According to this concept, to understand the tangible trace (of an event) at the empirical level, one needs to imagine the abstract world where the alleged event took place and a version where the event did not happen. The abstract trace represents the modification or difference between the event and non-event in the abstract world. In addition to being a presence, the trace can thus be the perceivable absence of something that previously existed but was obliterated by a previous event (Jaquet-Chiffelle & Casey, 2021, p. 2).

The authors underline that what is commonly referred to as a trace in scientific practice is, in fact, one or more *observable facets* of the tangible trace (Jaquet-Chiffelle & Casey, 2021, p. 2). Due to the limitations at the tangible level, the tangible trace can thus be partially perceived through the *observable facets* of the trace, which are never complete. The trace is mutable due to intrinsic or extrinsic events and can evolve (e.g., decompose) as time goes on, due to intrinsic events, even without extrinsic events. The authors emphasise that their conceptualisation of the trace is not a faithful representation of reality, nor a fact, but a concept representing the perceivable difference at a scene, resulting from an event of interest. The difference relates to a former state of the scene *modified* to the subsequent state, due to something being changed, added, or removed. The modification may be perceivable as a trace at an abstract or tangible level: A fallen tree in the woods is a perceivable trace of an event, in spite of no one seeing or hearing it fall, and magnetism happens even though we cannot see or feel it. The tangible digital trace is defined as “the modifications of the scene, subsequently perceptible in binary form, resulting from the event of interest and subsequent intrinsic events” (Jaquet-Chiffelle & Casey, 2021, p. 10). An observation instrument is always required in the digital realm, since data cannot be observed directly.

Jaquet-Chiffelle and Casey’s trace theory builds on Peirce’s sign chains and the three *interpretants*. Although they acknowledge that one trace facet can have different meanings, depending on the cohesive consideration of other facets and their context, the interpretants and effects on the interpreter (observer) are excluded from the paper’s theoretical discussion

(Jaquet-Chiffelle & Casey, 2021, pp. 1, 3). Interpretation is, however, a central concept in social science perspectives, and an STS perspective of the trace is presented below.

3.1.2 The reported trace as an actor with agency

The semiotic and quantitative descriptions of the trace do not elaborate on the function of the trace. Trace is sometimes associated with human traits/qualities, such as memory: “This is evidence that does not forget” (Kirk, 1953, p. 4), or being a testimonial witness: “The purpose of forensics – making mute things give testimony – implies a process of adding informational value through analysis in order to move from the things as occurrences in and of themselves to things as evidence to propositions” (Tilstone et al., 2013, p. 20). As discussed in Sunde (2020a, p. 22), some even state that “the mute witnesses never lie” (Arntzen, 2018, p. 9, my translation), which may create a notion of credibility of the trace and associated claims.

Instead of assigning human traits to the trace, the concept of non-human entities with agency may be helpful. This perspective is applied in several scholarly traditions, such as actor-network theory and critical realism. As pointed out by Kruse (2016, p. 94), the trace itself is often not present or in the criminal case, but the “written traces” are described in reports by forensic scientists or crime scene investigators. The written trace may not be just text – it can be represented and visualised in various ways, such as tables, graphs, drawings, and photos. The written trace becomes a stable inscription, which can be moved around between actors in the justice system. However, the written representation may differ from the observed facets, due to the subjectivity involved in the investigation and documentation process. Kruse (2016, pp. 110-112) highlights that the document may be stable, but the reported knowledge objects can still be interpreted in various ways, for example, by different epistemic cultures. From this perspective, the written representation of the trace is prone to different interpretations by the actors involved further in the investigation process, such as the legal decision makers. The interpretative flexibility and possible implications of this is discussed in **Article 3** (see section 3.2.3). **Article 5** sheds light on how such activities were performed and documented and provides insight into how the notion of credibility may be mediated, by both information presence and absence.

3.2 From trace to storyline – connecting the dots and crafting a scenario

The thesis’s research question centres on the DF practitioner’s role in the investigation process, from trace to evidence. Casey (2013) states that DF is neither a scientific inquiry nor an investigative tool – it is both. Of great importance is how the DF practitioner understands

their own role, including their own expertise and limitations. How the DF practitioner perceives the inquiry is also crucial, such as whether they believe to be performing a scientific inquiry or non-scientific investigation, and whether they understand their own capabilities to apply the scientifically derived and proven methods and to apply rigorous scientific reasoning to underpin any truth claims. Still, the value of digital evidence is assessed and decided by legal decision makers. Central here are perceptions of beliefs about the scientific maturity or “scientificness” (M. Olivier & Gruner, 2013, p. 34; Shaw, 2001, pp. 656-657) of DF. M.S Olivier (2016a, p. 47) argues that “if forensic science is not scientific but a pretence of science, trust in the endeavour is misplaced”. An inflated belief in scientificness may lead to overconfidence about the claims of truth and less scrutiny of the DF process and its outcome.

As discussed in section 2.3, there is little empirical research on how DF investigations are actually performed, and there is thus limited knowledge about the DF practices reflection of scientific maturity. Section 3.2.1. discusses various aspects of scientificness that were essential when looking for traces of applied science in the DF reports. The aspects concerning scientificness inspired the analysis in **Article 4**, which examined the hypotheses’ role for the inquiry and the techniques applied for safeguarding examiner objectivity and evidence reliability, and **Article 5**, which explored the documentation and reporting of DF procedures and results. Performing an investigation is not a concept on which there is unanimous agreement, and several typologies such as science, art, craft, or the reflexive investigator have been developed. These typologies are outlined briefly, since they aid the summary discussion in section 6 about the characteristics of the observed practices in the empirical foundation obtained through the DF experiment.

The thesis’s objective was, however, not to reduce the work of DF practitioners to a category or typology but to gain deeper insights into what they actually do when analysing an evidence file and reporting the results. This inspired a shift in the analytical perspective, and **Article 3** draws on central concepts from STS to understand how the DF practitioner transforms the observations of traces into written representations in analysis reports aimed for use as digital evidence in a legal context. The theoretical concepts, such as interpretative flexibility, narrative construction, and inscription, are elaborated in section 3.2.2.

3.2.1 DF casework – as scientific inquiry or investigation

Since DF is “the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence [...]” (Årnes, 2018, p. 4 referring to Palmer, 2001, p. 16), the section explores three dimensions related to the *scientificness* of DF inquiry during the analysis and presentation stages of the DF process:

- the scientific underpinning of the DF discipline (is DF a science?)
- the scientific expertise of the practitioner (is the DF practitioner a scientist?)
- the enactment of science (is DF examination a scientific inquiry?)

These above-mentioned scientificness aspects are interconnected and may underpin truth claims in various ways. Nonetheless, *illusions* of scientificness can foster unjustified assumptions of truth. The correct application of scientifically derived and validated methods can be a scientific inquiry – but it does not make the practitioner performing the procedure a scientist. Being a scientist would entail competence in performing scientific inquiry but not a guarantee that a scientific procedure has been followed. The methods, tools, and procedures applied by scientists may have a scientific underpinning, but that does not validate the discipline, in which they are applied, as a science.

Scientific underpinning

The scientific status of DF has been and still is debated (e.g., N. Beebe, 2009; Carrier, 2006; Casey, 2020a, 2020b; Cloppert, 2009; Cohen et al., 2011; Gladyshev, 2004; Losavio et al., 2016; M. S. Olivier, 2016a, 2016b). For many jurisdictions, such as the USA and the UK, the scientific underpinning is essential for the admissibility of the forensic evidence in a court of law (e.g., Pollitt, 1995). Other jurisdictions, such as Norway and Sweden, do not have a similar formalised admissibility threshold. Independent of admissibility, the scientificness of the DF discipline has been subject to scrutiny from a philosophical stance, discussing whether DF is founded on a robust theoretical underpinning (Tewelde et al., 2015).

The DF discipline originates mainly from the computer science domain. There are different opinions about the epistemological status of computer science, and the discussion revolves around whether it is a mathematical, engineering, or scientific discipline (Angius et al., 2021,

8.1-8.3). Computer science centres on computational systems and problems such as hardware, software, abstraction level issues, artefacts, programming, algorithms, physical computation, and verification/testing/experimentation methodology (Angius et al., 2021). The scientific underpinning of computer science is of great importance to exploring questions of digital events and behaviours of computer systems. However, since DF is concerned with investigating incidents involving human actors and human-computer interactions, the computer science foundation provides an inadequate underpinning for this dimension.

Scholars from the computer science domain, such as Gladyshev (2004), Carrier (2006), and Cohen (2013), have provided substantial contributions to the theoretical underpinning of DF as a science. Gladyshev (2004) aimed to make event reconstruction in DF science more rigorous and objective by introducing mathematics in DF analysis. Carrier (2006) aimed to define a theoretical model of a computer's history based on the theory of computers. Central to both Gladyshev's and Carrier's research on event reconstruction were finite state machines, which are abstract machines that can be in exactly one of a finite number of states at any given time. The finite state machines were also discussed by Cohen (2013) but were not central to the theoretical development in his research. Cohen's objective was to strengthen DF's theoretical scientific underpinning and proposed a formalised model for depicting the inherent nature of DF trace evidence in the legal context. The *trace* concept was advanced in the above-mentioned paper by Jaquet-Chiffelle and Casey (2021, see Section 3.1.1), using mathematical language. Experimentation, testing, and validation have been subject to several theoretical contributions, such as the ExperDF-CM conceptual model for DF experimentation (Oliveira Jr et al., 2020), the Framework for Reliable Experimental Design (FRED) (Horsman, 2018), the Digital Evidence Reporting and Decision Support framework by Horsman (2019), and the model for reliability validation of file system interpretation (Nordvik et al., 2021). Academics in the field have also provided valuable theoretical contributions to a formal evaluation of digital evidence (e.g., Casey, 2020b; Ryser et al., 2020; Tart, 2020). Although DF rests on the robust underpinning of computer science, as an applied scientific discipline, it seems to be under construction, with a less robust scientific foundation through long-term development, testing, refining, and legal and scholarly scrutiny, compared to other applied forensic science disciplines such as the DNA discipline (M. S. Olivier, 2016a, 2016b).

Scientific expertise of the practitioner

M. S. Olivier (2016a, p. 49) draws an important distinction between being a scientist and being involved in the forensic science process. The process itself (and, hence, its development) needs to be scientifically sound; the scientific laws that underlie the process need to be understood (and justified) by the developers of such a process. Still, the many people involved daily in a forensic science process are not (and need not be) scientists, and conducting the process does not make you a scientist. While a non-scientist DF practitioner may safely perform scientifically validated procedures and describe the output, the scientist's expertise is crucial for interpreting results. Interpretation is involved when determining what traces are and what they mean, evaluating the value of the trace in the context of the questioned matter, as well as identifying, understanding, and explaining the limitations and uncertainties associated with the results. These are aspects of the expertise challenge described in section 2.2.2. Although many academic institutions offer DF education, there are currently no generally accepted standard minimum requirements regarding who is an expert in the DF field throughout the world. The consequence is a number of formally qualified experts, all having different levels of competence (Humphries, 2019, p. 40; Watson & Jones, 2013, p. 826). In her PhD thesis, Humphries (2019, p. 98) highlights an important point about learning the necessary skills to perform DF examinations: most of the courses aimed at DF in the UK sit in computer science departments, where the primary focus is computer science, with a few modules directed towards broader forensic elements. Qualification in scientific inquiry entails competency in applying the scientific method to research problems aimed to produce generalised knowledge. Still, this competency may not provide sufficient qualification to perform what is required in a DF investigation, which entails applying science in a forensic examination to produce knowledge about a specific event. One of Humphries' informants stated that "placing a forensics course in a computer science department is the reason many fail as they produce computer scientist not forensic experts and the gap is vast!" (Humphries, 2019, p. 98).

Being a scientist within the applied DF science brings about different challenges from those of disciplines with a more static problem to investigate. DF is concerned with a rapidly changing technological environment, complicated by the human-computer interaction dimension of usage patterns and attempts to hide, obfuscate, or destroy traces. The formal expertise may thus be rapidly outdated, and continuous training and proficiency testing is necessary to provide objective evidence of sufficient expertise compared to some pre-

established criteria (Garrett, 2021, pp. 106-107). To assess whether the DF practitioner had sufficient – and an adequate type of – expertise for the DF examination, documentation about their expertise is vital.

The enactment of scientific inquiry vs a DF investigation

S. H. James and Nordby (2002, p. 6) highlight the difference between the underlying science and the applied forensic science: While science typically is developed in a controlled environment, forensic science usually applies scientific theories and methods in chaotic environments. And, while forensic science typically is about reconstructing past events, science is usually concerned with predicting what will happen in the future. Reconstructing activities based on artefacts is often a demanding task, as it is easier to predict what a computer program will do than to reconstruct what it did (M. S. Olivier, 2016a, p. 51). M. S. Olivier (2016a, p. 48) underlines a vital distinction, “If forensic science is the use of science to help answer disputes in legal and related matters a question that arises is when this science is actually performed”. Confusing investigative improvised methods with scientific rigour may lead to misinterpretations about the credibility and evidential value of the results.

In contrast to the attempts to justify that DF is a scientific discipline, more academics have been concerned with exploring how DF investigation *ought to be* performed to demonstrate scientificness during the inquiry. The OSAC (The Organization of Scientific Area Committees for Forensic Science)/NIST subcommittee for digital and multimedia evidence addressed the issue of whether the digital/multimedia discipline can demonstrate that the processes, activities, and techniques used are sufficiently *scientific*. As a result, they produced a harmonising framework for forensic science practices and digital/multimedia evidence (Pollitt et al., 2018, p. ii). They state: “to be scientific, a discipline must employ scientific reasoning” (Pollitt et al., 2018, p. 3) and outline the scientific reasoning process as applying abductive, deductive, and inductive reasoning, sometimes referred to as the hypothetico-deductive model. The paper defines and describes core forensic processes (authentication, identification, classification, reconstruction, and evaluation) and a set of forensic activities (survey, analysis, integration, interpretation, and documentation) for producing information to feed into the forensic processes (Pollitt et al., 2018, pp. 6, 11). The core forensic processes and activities interact with each other in various ways to fulfil the objective of the scientific inquiry.

The core forensic process of evaluation is central in the Case Assessment and Interpretation (CAI) framework, which was developed as a novel application for forensic science to deal with forensic science problems and opinion evidence (R. Cook et al., 1998a, 1998b). CAI was founded on Bayes' Theorem's logical framework (Bayes, 1764), which was applied to deal with uncertainty through subjective probabilities (Jackson, 2011, p. 2). Perhaps the single most fundamental element of the CAI model is the definition of a *pair of propositions* from which likelihood ratios for scientific findings can be derived (Jackson, 2011, p. 9). The hierarchy of issues (originally named the hierarchy of propositions) is a central component of the CAI framework. It aims to formalise and clarify the precise contribution of forensic science evidence in a particular case (R. Cook et al., 1998a, 1998b). It separates the expert opinions at source, activity, and offence levels. The source level deals with propositions about the source of the questioned material. The activity level is concerned with the activity the suspect allegedly has done, incorporating the source level issues. The offence level relates to the allegedly committed crime and is concerned with both the legal application of the phenomenon and the issues of criminal guilt. As with the activity level, it incorporates the levels below. The forensic science experts would typically be commissioned to provide source level opinions and sometimes activity level opinions. However, offence level opinions are regarded as being outside the forensic scientists' area of expertise and should be the task of the legal decision makers (R. Cook et al., 1998a, p. 233). The CAI framework distinguishes between the forensic scientists' role at different stages of a criminal investigation and which expert opinions they may provide. According to Jackson et al. (2006, p. 39), the forensic expert may provide three types of opinions. The first type is "investigative opinions", which are explanations, or conjectures, for observations, sometimes associated with posterior probabilities for the explanations. The second is "preliminary evaluative opinions", which are expressions of the likelihoods for the findings, given the truth of individual propositions. The third type is "fully evaluative opinions", which are numerical or verbal expressions of the magnitude of the likelihood ratio. Whilst investigative and preliminary evaluative opinions are more relevant at the investigative stage, fully evaluative opinions aim to guide the legal decision makers at the court stage. Guidelines for DF work also refer to the CAI framework when stating how evaluative opinions should be structured (ENFSI, 2015a, pp. 34, 41; 2015b).

According to M. S. Olivier (2016a, p. 48), the discourse in the DF domain has only seen limited self-reflection about the use of science (or scientific methods) in its activities, and,

while there are some exceptions, the few published claims that DF is scientific are often based on a limited understanding of science.

The enactment of (non-scientific) investigation

As mentioned above, DF may be both a scientific inquiry and an investigative tool. However, suppose the examination of digital traces fails to fulfil the requirements for being a forensic inquiry and, instead, is performed as a non-scientific investigative inquiry. In that case, some central questions are: how should the results be understood, and can and should they be trusted?

An investigation is typically defined as being much broader than a scientific inquiry. Although a criminal investigation aims to find the truth and uses research-based methods and tools, the investigation stage may involve more ad hoc and improvised methods and tools (Hewling, 2013, p. 199; Ward, 2021, pp. 103-104). The function of investigative hypotheses is to support the investigation of a particular event. Thus, they deviate from the function of scientific hypotheses, which aim to develop a generalised theory (Tewelde et al., 2015, p. 38). Due to the nature of a criminal investigation, the hypotheses are often derived through abductive reasoning based on uncertain and incomplete case information, combined with experience and formal knowledge of the practitioner (Rønn, 2013, p. 281; Sunde, 2020b, p. 3). The early stages of an investigation aim to uncover potentially relevant information to answer questions about what happened, who was involved, how it was performed, why, where, and when (T. Cook, 2016, p. 38). Without formal evaluation, the investigation can produce descriptions of the uncovered traces, interpretations of what the traces mean, and inferred explanations related to the investigative questions that have been explored (Jackson et al., 2006, p. 39). Lipton (1991, pp. 60-61) suggested that there are two epistemic filters involved in the abductive investigation process. The first filter reduces all possible explanations to the plausible explanations. The second filter reduces plausible explanations to the best explanation.

Research from the social sciences has shown that investigation may be understood as different types of detective work based on its characteristics and produced typologies such as art, craft, or science (Tong & Bowling, 2006). In short, the *craft* typology emerged from learning by doing and on-the-job experience with investigation and case management. Essential for this perspective is learning to use the tools to build a case (Hald & Rønn, 2013, p. 25; Tong & Bowling, 2006, p. 324). Detective work as *art* is characterised by using instinct, hunches, and

intuition for solving investigative problems (Tong & Bowling, 2006, p. 324). From this perspective, the detective is an artist with inherent skills that only a few talented people possess (Hald & Rønn, 2013, p. 26). According to the *science* typology, the investigative work is related to or conditioned by science and is closely related to the scientificness perspective discussed earlier in this section. The detectives are skilled in scientifically founded approaches, such as crime scene management, handling physical evidence (Tong & Bowling, 2006, p. 325), investigative interviewing (e.g. Gabbert et al., 2018; Griffiths and Rachlew, 2018; Jakobsen, 2021) and investigative decision-making (Fahsing, 2016). Hald and Rønn (2013, p. 28) highlight that this perspective relates closely to the view that investigation is or aims to be an evidence-based enterprise, which entails that there is an empirical underpinning that the applied methods or activities are effective.

Hald and Rønn (2013, pp. 30-34) suggested adding the *reflexive* investigator (detective) typology with critical, methodological, and analytic awareness and expertise for the methodology of discovery. Interpretation plays an essential role in the reflexive detective's work and is based partly on background knowledge and partly on the traces that exist in the individual case. The reflexive investigator can reflect critically upon their own investigative practice and determine the robustness of assessments and conclusions (Hald & Rønn, 2013, p. 31). These perspectives are helpful for theorising DF inquiry and practice in a more nuanced way than the dichotomy of science or investigation, and is further discussed in section 6. They also pave the way for applying other theoretical concepts from social science, such as interpretative flexibility, narrative construction, and inscription, to explore in depth what DF practitioners actually do when analysing and reporting digital evidence.

3.2.2 DF casework – as interpretation, narrative construction, and inscription

The situation created by the DF experiment enabled a novel and multifaceted exploration of investigative practices, which otherwise would be black-boxed. **Article 3** explores the different interpretations of the same artefact or trace, which adds to the concept of evidence dynamics and sheds new light on whether digital evidence should be viewed as objective and reliable. **Article 3** draws on central concepts from STS research, such as interpretative flexibility, narrative construction, and inscription, to understand how the DF practitioner transforms the observations of traces into written representations in analysis reports aimed for use as digital evidence in a legal context. The central theoretical concepts are therefore elaborated.

Narrative construction and interpretative flexibility

Narrative construction is a useful perspective for understanding the process of establishing knowledge about the particular, as opposed to developing generalised knowledge – which is the aim of scientific research. Narratives connect people and their actions to the crime (Kruse, 2016, p. 19). In a legal context, the narrative helps assess the value of individual traces and the traces in context. Kruse (2016, p. 32) highlights that, in contrast to, for example, witness accounts, forensic evidence does not come in narratives. It is regarded as reliable and often becomes the anchoring point in the narratives of the case (Kruse, 2016, pp. 33, 156).

Nevertheless, forensic evidence is unable to stand on its own and, thus, a context is necessary to evaluate its value to the case under investigation (Kruse, 2016, p. 33). In contrast to verbal evidence, forensic evidence cannot present itself, and a human must do the representing (Kruse, 2016, p. 10, referring to Barad, 1981).

The physical forensic evidence has primarily been related to source level issues (see hierarchy of issues, section 3.2.1) and, more rarely, to activity and offence levels (Kruse, 2016, p. 79). Whether a scientist should address the activity level issues in formal evaluations is contested, since it is often necessary to obtain and consider case-specific information outside the scientist's expertise and the fact that interpretation at the activity level will be conditioned by the aspect of time (Evetts et al., 2000, pp. 7-9; Risinger, 2013, p. 70). For example, DNA evidence may be found on the knife, but the trace itself says nothing about the event that led to the DNA trace on the knife. At this point, digital evidence differs considerably from the other forensic disciplines. Events, and thus activity level issues, are often the primary focus of the DF investigation. An event is described as “a complete collection of related things that have happened (or are happening) in a world within a specific closed interval of time” (Jaquet-Chiffelle & Casey, 2021, p. 4). Establishing who caused the event in a digital environment is typically a difficult task – and must often be supported by other investigative steps, such as suspect interviews. The event can be linked to information indicating who may have caused the event, but association with a high/strong probability, such as a fingerprint or DNA, is often out of reach. Also, separating human-related activities from system-generated activities may also be difficult, due to the computer systems' complex and diverse nature.

Interpretative flexibility (Collins, 1981, p. 4; Doherty et al., 2006, p. 569) is a concept central to understanding the variance observed in the DF experiment and the DF practitioner's role in constructing what starts as a trace and becomes investigative leads or digital evidence.

Interpretative flexibility is a central concept of social constructivism in STS and is often

applied to explore social negotiations between different groups about disputed scientific findings or controversies about technology (Silvast & Foulds, 2022, pp.109-110). Interpretation and subjectivity are also acknowledged within the natural/computer science perspectives. Yet, from the conservative stance, it is understood as something that can and should be minimised or avoided. The thesis applies the general description of the concept, namely that concepts have flexibility when they are interpreted differently (Silvast & Foulds, 2022, p. 110) and explores the function of interpretative flexibility in the interpretation of individual traces and in the narrative construction. The thesis adds to the work of Kruse (2016), Santos (2014), and Dahl and Sætnan (2009), who have all explored the interpretative flexibility of various types of forensic evidence. It also adds to the insights provided by Brookman and colleagues (2021; 2020a;), who examined the role of digital evidence (including CCTV and phone data) in homicide investigation narratives.

Inscription, mediation, and epistemic distance

Reporting and documentation practices are fundamental to the construction of the trace as a knowledge object and how legal decision makers perceive its evidential value. Although not adopting the full theoretical foundation, the thesis is inspired by actor-network theory (ANT). **Article 3**, in particular, draws on central ANT concepts, to explore inscription practices and the role of DF practitioners as obligatory passage points and mediators of digital evidence, and to scrutinise and nuance the myth of a DF process characterised as mechanical objectivity and digital evidence as mere *facts*. In addition to the centrality of these concepts for **Article 3**, the ANT concepts are relevant for a holistic analysis of the documentation practices referred to in **Articles 3, 4, and 5** – and are therefore applied to the concluding analysis and discussion in section 6. The applied concepts are outlined below.

ANT was developed in the 1980s by Bruno Latour, John Law, and Michel Callon (Skjølvold, 2015, p. 24). Central to ANT is the symmetry between humans and non-humans or things as equal actors (referred to as actants) involved in information processing and knowledge production (Skjølvold, 2015, p. 67). The symmetry is a methodological choice which facilitates the empirical study of the different modalities of agency (Callon, 2001, p. 65). The actants are tied together into networks built and maintained to produce the power necessary to achieve a particular goal (Skjølvold, 2015, pp. 25, 77). In contrast to the more traditional perception of a network as a technological structure of interconnected nodes, in ANT, it is a concept for understanding and examining relations and roles, where human and non-human

actors interact and influence each other in what are referred to as “actor-networks” (Skjølsvold, 2015, pp. 68, 78). According to the symmetrical view, both human and non-human actants may be *intermediaries* or *mediators*. An intermediary translates interest and transports meaning without transformation (Latour, 2005, p. 39). The interest can be in the form of, for example, a text, a product, a service, or money, which in the thesis primarily concerns the DF practitioners’ reported results in analysis reports. In contrast, the mediators “transform, translate, distort, and modify the meaning of the elements they are supposed to carry” (Latour, 2005, p. 39) and have faculty to make other entities do something different from expected (Latour, 2005, pp. 58-59). Actors with the power to define and control may become “obligatory passage points”, which Callon (1986) describes as central and highly skilled actors within the particular field of expertise. The concepts were applied in **Article 3**, when exploring factors that empower the DF practitioners to mediate the digital evidence.

DF practitioners transform digital traces into evidence by *inscribing* them into reports. The concept of “inscription”, which was first used by Latour and Woolgar (1979/1986) in the study of knowledge production in scientific laboratories, is central for the thesis to understand how the traces become digital evidence prescribed with a value in DF reports. Inscription is understood as a proliferation of words and things that can take many forms, such as photos, maps, graphs, diagrams, films, acoustic or electric recordings, observations noted in a laboratory logbook, and illustrations (Callon, 2001, p. 62). To describe how the forensic report becomes a mediator, Santos (2014) uses the term “Epistemic distancing” which is outlined as “a professional *ethos* marked by distinctions and differentiations from the language, practices, classifications, hypothesis and opinions of the police” (p. 200, italics in original). A vital means in epistemic distancing is “interpretative limitation” (Santos, 2014, pp. 191,193), which refers to the discrepancy between the questions the investigation seeks answers to and what the laboratories agree to provide answers to, as well as the discrepancy between the type of answers the police desire (categorical yes/no) and what the lab offers (probabilistic evaluation). The concept of inscription was applied in **Article 3** to theorise how the DF reports may intentionally or unintentionally be crafted to mediate the perception of evidential value by what is included or excluded and how the traces and the related context are described.

3.3 Evidential value of traces

Due to their relevance to the overall research question, aspects concerning the evidential value of digital traces are discussed from different angles across all articles. As described in section 3.1.1, Jaquet-Chiffelle and Casey (2021, p. 7) conceptualise the trace as the perceptible difference between an event and a non-event in the abstract world. The trace consists of several “facets” (Jaquet-Chiffelle & Casey, 2021, p. 2) that can be observed at the physical, binary, application, or semantic levels, which are used for classifying the trace, i.e., establishing what the trace is. Still, the trace’s *evidential value* is not determined merely by observing it. Instead, it is constructed by relating the trace (or the absence of a trace) to an evidential theme, that is, revealing the occurrence or non-occurrence of an event (Anderson et al., 2005, p. 74). When the DF process is performed in the context of a criminal investigation, its aim is to obtain traces and transform them into meaningful evidence in a legal context. The thesis draws on the conceptualisation of evidential value presented by Anderson et al., which is rooted in a rationalist stance (2005, pp. 63-67, 81). The theory about evidential value builds on fundamental assumptions about legal proof: that knowledge about past events is possible, and the establishment of the truth of alleged facts in adjudication is typically a matter of probability, where absolute certainty is often out of reach (Anderson et al., 2005, p. 82). The concepts are applied here to enable an epistemic dialogue between the scientific criteria for truth claims and the legal assessments, inferences, and decisions about guilt or innocence for a suspect or defendant.

3.3.1 Demonstrative tangible evidence

Anderson et al. (2005, pp. 63-67) distinguish between testimonial and tangible evidence. Further, the tangible category may be separated into *real tangible evidence*, which is a thing itself, or *demonstrative tangible evidence*, which concerns not the thing itself but representations or illustrations of such things (Lempert et al., 2000, pp. 1146-48; Tecuci et al., 2016, p. 120). Digital traces are not observed directly but through technology. They are not tangible and thus need to be represented through descriptions, visualisations, and illustrations of files; they therefore fall into the demonstrative tangible category. Demonstrative tangible evidence has three primary credentials: relevance, credibility (believability), and inferential or probative force/weight (Anderson et al., 2005, p. 60; Tecuci et al., 2016 p. 62).

Relevance is a trait that makes the evidential theme (the matter to be proved) more or less probable. The relevance may be either “direct”, as directly linked to the matter to be proved,

or “indirect”, by being evidence about other evidence, for example, the credibility of a witness (Tecuci et al., 2016, p. 63). The relevance of an item of information often depends on what other items of information have been obtained (Tecuci et al., 2016, p. 63).

The *credibility* of tangible demonstrative evidence refers to three elements: Authenticity, accuracy/sensitivity, and reliability (Anderson et al., 2005, pp. 64-65). To assess the *authenticity*, one considers whether the evidence is a genuine representation of what it appears to be (Anderson et al., 2005, pp. 64-65; Tecuci et al., 2016, p. 120). Since digital evidence is prone to intentional or accidental manipulation, authentication is vital. For example, CCTV timestamps must be verified, to ensure that the recording corresponds to the actual point in time when the crime occurred. An assessment of the *accuracy/sensitivity* of evidence is concerned with whether it provides a sufficient resolution to discriminate between possible events/explanations (Anderson et al., 2005, p. 65; Tecuci et al., 2016, p. 121). For example, a too low resolution on CCTV footage increases the risk of erroneous identification. An assessment of the *reliability* concerns whether the evidence was produced in a repeatable, dependable, and consistent manner (Anderson et al., 2005, p. 65; Tecuci et al., 2016, p. 121). Anderson et al. (2005, p. 65) relate reliability to the operating characteristics of the device used to generate it. In a DF context, multiple devices must be taken into account when establishing reliability. The device involved in generating the trace must be scrutinised, as well as the devices and technology used for securing and examining it in the DF context. For example, in an alleged crime of downloading child sexual exploitation and abuse material, the reliability is concerned with the device used for performing the activity and generating traces in the first place, and secondly, with the analysis hardware and software used to examine the traces during the DF investigation. Error mitigation and verification are necessary to safeguard the reliability, and **Articles 4 and 5** shed light on whether such activities were performed and documented during the DF experiment. However, the thesis highlights another aspect of the reliability dimension, namely, the reliability of the *human instrument* that turns the digital trace into demonstrative tangible evidence, which is primarily explored in **Articles 2 and 3**.

Safeguarding evidence credibility relates directly to the principle of the right to a fair trial and the presumption of innocence (Stoykova, 2021). Yet, several procedural factors in the DF process may undermine these minimum legal safeguards. Stoykova highlights several issues as threats to the fairness of DF investigations, such as overreliance on and inappropriate use of investigative technology, inadequacy of the defendant’s opportunity to challenge or cross-

examine the dataset for exculpatory evidence at the pre-trial stage, and inadequate reliability testing in DF practices.

Inferential/probative force or weight of evidence is about how strongly the evidence favours or disfavors particular hypotheses or propositions and is described in probabilistic terms (Tecuci et al., 2016, p. 67). Probabilistic judgements can be expressed numerically or verbally, a matter that has been subject to debate in the forensic science community (see, e.g., Arscott et al., 2017; Martire et al., 2014; W. C. Thompson & Newman, 2015). However, since all probabilities rest upon arguments, the probability is more about structuring arguments than about numbers – and if the arguments are faulty, the determined probabilities will make no sense (Tecuci et al., 2016, p. 67, referring to Shafer, 1998, pp. 5-9). **Article 5** focused particularly on how the conclusions were articulated in DF work and how (un)certainty descriptors were used to describe the evidential value.

3.3.2 Testimonial evidence

Although not investigated by the thesis, it should be noted that, since the digital evidence is presented orally in court, the value of digital evidence may also relate – implicitly or explicitly – to the credentials for testimonial evidence. There are two basic sources of uncertainty related to testimonial evidence: competence and credibility (Tecuci et al., 2016, p. 122). Competence relates to whether one has had access to the reported information and has the knowledge, skills, and professional experience to understand and interpret the information (Tecuci et al., 2016, pp. 122-123). For an expert witness presenting opinion evidence, the competence is critical, since the expert not only describes traces to the court but also presents opinions concerning the force or weight of the evidence. The validity of an expert opinion relates to “foundational validity”, which requires that the test or method is scientifically sound, and “applied validity”, which considers the merits and limitations of the methods and tests when applied to a particular piece of evidence (President's Council of Advisors on Science & Technology, 2016, p. 43). The validity of an expert opinion may also relate to “evaluative validity”, which entails that “the expert’s opinion is transparently rooted in empirical data or studies and appropriately insulated from prejudicial information or other sources of cognitive bias” (Carr et al., 2020, p. 4, Fig. 2).

3.4 Error and uncertainty

The fair administration of justice rests on the justice system's ability to base its legal decisions on sound or true knowledge, and the thesis's research question addresses the DF practitioner's role in mitigating error and misinformation in the context of the DF process and criminal investigation. The concept of *error* recurs across all included articles and the thesis summary discussion. The notion of error is thus insolubly related to what is true or correct. With reference to John Stuart Mill, Hon (1995, p. 8) argues that we can never know what a thing is unless we are able to give an adequate account of its opposite. Categorising something as an error implicitly entails a deviation, but from what it deviates depends on the philosophical perspective of what constitutes truth or correctness. Finding a generally accepted description of the phenomenon of error has been one of philosophy's very serious and crucial problems (Hon, 1995, p. 5). This section limits the discussion to perspectives relevant to DF, and errors related to the analysis and presentation of digital traces. The categories of error applied in the thesis are described in section 3.4.1. Since the DF practitioner's role in error mitigation is the centre of attention in the thesis, the central applied concepts concerning cognitive and human factors are discussed in section 3.4.2.

3.4.1 Categorisation of error

Error in DF investigations may be divided into two general categories: technical and practitioner error. The individual cognitive and human factors constitute the focal point of the thesis and the baseline in all included articles. When discussing human error, the thesis centres on unintentional practitioner error and includes blunders, slips, lapses, and mistakes, when referring to error. However, from a quality management perspective, failing to detect and correct implementation error may also be considered a practitioner error, and these categories are outlined below.

The technical error category encompasses *error in technique* and *error in the implementation of techniques in tools* (SWGDE, 2018, pp. 9-11). In terms of the first category, techniques are the basis for processing data for different purposes, such as copying data, creating a cryptographic checksum (hashing), searching for data, and recovering deleted files. During such processing, random errors may occur, and the techniques can sometimes be characterised with an error rate (SWGDE, 2018, p. 10). In contrast, error in the implementation of techniques in tools leads to systematic error. The flaws are triggered by particular conditions

that result in an incorrect output, which is incomplete, inaccurate, or misinterpreted (SWGDE, 2018, pp. 3-4). Since they are systematic, they will reproduce the error every time the particular conditions occur. For example, it was discovered that a version of the software Cellebrite inaccurately interpreted timestamps from iOS phones with Apple File System (APFS), which led to a review of several criminal cases, to assess whether the misinformation from the erroneous output had led to errors of justice (Grut, 2020).

Practitioner error refers to a mistake or an operator (human) error. It may be random or systematic, related to negligence or incompetence and is, for the most part, unintentional and unquantifiable (Christensen et al., 2014, p. 124). Practitioner error can relate to physical tasks, such as handling the digital device and preserving its state. Inadequate handling of digital devices or spaces with relevant traces to the case under investigation may lead to alterations of the trace. Such alterations are referred to in DF literature as “evidence dynamics” and described as “any influence that changes, relocates, obscures, or obliterates evidence, regardless of intent between the time evidence is transferred and the time the case is resolved” (Casey, 2011b, p. 27). Practitioner error may also relate to cognitive tasks, such as perception, interpretation, inferences, decisions, and conclusions (see section 3.4.2). These errors do not cause evidence dynamics but occur in how the evidence is described, visualised, or represented as a knowledge object, and thus how others perceive it. Practitioner error is difficult to estimate but can be mitigated through quality assurance systems, training, proficiency testing, peer review, and adhering to validated protocols and discipline best practices (Christensen et al., 2014, p. 124).

Since DF investigation is not only a scientific inquiry but also an investigation, it could be argued that errors related to law and ethics should be added to the practitioner error category. When DF investigation is performed at the pre-trial stage, it operates under the criminal procedure regulations of the jurisdiction. An error in this sense is thus a deviation from legal obligations, such as safeguarding the presumption of innocence – or what is legally accepted, such as going beyond regulations for search and seizure. However, the court can also make erroneous decisions, as either false positives (arrests and convictions of the innocent) or false negatives (failure to arrest and convict culpable offenders), often referred to as miscarriages of justice (Bushway & Forst, 2013, p. 211).

Hon (1995, p. 6) distinguishes between two ways of going wrong – through mistake and error. She associates making a mistake with avoidable ignorance and error, where the mistake could have been avoided by checking known and available procedures. An error is associated with unavoidable ignorance and happens since the phenomenon is novel and lacks a well-studied and agreed-on standard procedure. Practitioner error may be intentional, such as fraudulent behaviour. These may be referred to as violations, which are “*intentional* actions or decisions not to follow procedures, rules or instructions” (Bridger, 2021, p. 25, italics in original). Practitioner errors may also be unintentional, resulting from blunders such as transposing numbers when recording data, incorrect instrument use, selection of inappropriate methods, or improper method application (Christensen et al., 2014, p. 124). These may either be *slips and lapses* (execution failures), where there is a discrepancy between the intended action and what was actually done, or *mistakes* (planning failures), which involve a mismatch between the prior intention and the intended or planned consequences (Reason, 1990, p. 8).

3.4.2 Cognitive architecture and mechanisms as sources of error

The DF practitioner is the focal point of the thesis’s research. The cognitive and human factor’s role in constructing digital evidence and mitigating error was explored from a theoretical angle in **Article 1**, focusing mainly on the concept of bias and the biasing sources. **Article 2** explored contextual bias and reliability between DF practitioners, with reference to the Hierarchy of Expert Performance (HEP) and advances the insights concerning expert decision-making in relation to four HEP levels. The concepts of bias, biasing sources, and reliability are discussed below.

Human factors in DF decision-making

Despite procedures aiming at objective analysis of forensic evidence, the largely subjective human judgement is heavily relied upon during the observations, interpretations, and conclusions (Venville, 2015). Within the DF domain, there has been a shift from perceiving tools and technology as the primary instruments in the DF process towards a greater acknowledgement of the importance of the human factor for examining digital evidence (e.g., Cervantes Mori et al., 2021; Ferguson et al., 2020; Pollitt et al., 2018, p. 3). Cognitive psychology explores the internal mental processes of the brain. It therefore provides relevant theoretical perspectives for exploring the role of DF practitioners in the DF process and how they influence the result. In the psychological sense, *error* is “all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and

when these failures cannot be attributed to the intervention of some chance agency” (Reason, 1990, p. 9). Reason offers three classifications of human error: behavioural level (classifies error according to the observable features of erroneous behaviour), contextual level (focuses on causality, and draws attention to the local triggering factors (situational) and underlying error tendencies), and conceptual level (rests on assumptions about the cognitive mechanisms involved in error production) (pp. 11-12).

Reason (1990, pp. 3-4) deviates between variable and constant errors. Constant errors are predictable and much more manageable than variable errors, given that the factors that led to the errors are understood. The prediction involves the conditions under which the error will occur and the particular form it will take. The thesis, and particularly **Articles 1 and 2**, advances the insights of the variable and constant errors at the contextual level, in a DF investigative context.

Cognitive bias

A part of this scientific domain is concerned with systematic and predictable deviations from rational judgement or decision-making, namely the *cognitive biases* (Blanco, 2017, p. 1).

Many biases have been identified through research (see, e.g., Manoogian & Benson, 2018). These biases generally occur subconsciously and are largely uncontrollable (Nickerson, 1998, p. 175; Pohl, 2022, p. 7). The biases arise from errors in cognitive processing known as *heuristics*, which are shortcuts in reasoning that sometimes can lead to systematic, predictable, and directional errors in the process of decision-making (Kahneman et al., 2021, p. 151; Tversky & Kahneman, 1974, p. 1124). While heuristics can lead to errors in judgement, they are also a natural and necessary component of human processing, due to the vast amount of information we interact with daily. On one hand, they help simplify and categorise a complex world, while, on the other, the oversimplification can result in flawed categorisations and errors in judgement (Nickerson, 1998; Tversky & Kahneman, 1974, p. 1124).

The “confirmation bias” (Nickerson, 1998, p. 175), which is frequently mentioned in criminal investigation and forensic science contexts, relates to information observation and processing. It involves a tendency to search for information corresponding to our belief of what has happened and overlook and explain away information that contradicts our belief (Nickerson, 1998, p. 175). Over the last two decades, much attention has been devoted by researchers to bias in forensic science examinations (see an overview in Cooper & Meterko, 2019), which Kassin et al. (2013) coined “the forensic confirmation bias”.

Biasing sources

While bias relates closely to heuristics, external sources have also been shown to impact decision-making systematically. Dror (2020, pp. 7999-8002) divided the sources of bias into groups spanning from human nature to case-specific, which are discussed in detail in **Article 1**. The taxonomy was adjusted in a paper published after **Article 2**, where level 1 was changed from “Case Evidence” to “Data”, and “Personal Factors” was added at level 2 (Dror, 2020), but these adjustments do not affect the thesis’s findings and discussions.

The role of irrelevant contextual information in distorting observations and conclusions is referred to as *contextual bias*. The systematic literature reviews outlined above (Cooper & Meterko, 2019; Kukucka & Dror, 2022) showed that a solid and consistent research base substantiates that irrelevant contextual information may bias forensic decision-making. The thesis contributes to this knowledge by examining these issues within the DF domain. The empirical findings presented in **Article 2** indicates that the DF discipline is no exception and should consider contextual influences when developing error mitigation measures, similarly to other forensic disciplines.

Reliability

In the context of cognitive psychology, reliability is about making consistent observations and decisions. Research has shown that forensic experts are inconsistent in their judgements and decisions and also in their decision-making when making repetitive decisions (Dror, 2016). **Article 2** is the first to explore reliability in DF decision-making and suggests low reliability between DF practitioners at all examined levels.

Kahneman et al. (2021) use the term *noise* to describe variability in decision-making related to bias or reliability in their recent book, and devote a chapter to noise in forensic decisions. While bias leads to systematically skewed decisions, the sources and causal mechanisms may be uncovered and corrected. Noise is more problematic, since it does not appear systematically, and we do not know why the decisions are noisy (Kahneman et al., 2021, pp. 4, 90-93). Another challenge with noise is that it is inherently statistical. It becomes visible only when we think statistically about a collection of similar decisions (Kahneman et al., 2021, p. 219).

Whilst **Article 2** sheds light on the reliability from a statistical point of view, **Articles 3** and **4** advance the insights concerning this issue, by exploring the implications of low reliability through a qualitative lens. These papers explore how low reliability may manifest itself when

DF practitioners form hypotheses prior to the analysis, during the analysis of the evidence file, and when presenting the results in reports.

4. Methods

This section provides a detailed overview of the research design and methods applied in the thesis. The methods used in this study may be characterised as the consequence of an *emergent explanatory design* (Leedy & Ormrod, 2014, pp. 153, 270), with a fixed quantitative phase before a gradually more emergent qualitatively oriented phase. The collection and analysis of documents and survey data for **Articles 1** and **2** were part of the original plan, while the ideas for **Articles 3, 4, and 5** emerged when reviewing the rich empirical material collected during the DF experiment and related surveys.

The section is structured as follows: first, the sample, data collection, and material for the thesis are described, followed by an elaboration of the applied analytical procedures. Then, the research quality and ethical considerations are outlined and debated. Finally, my professional position and scientific worldview are described and discussed.

4.1 Sample, data collection, and material

Submission forms and descriptions of commissioning procedures for DF work were obtained for a richer background to the design of an ecologically valid experiment and are described in section 4.1.1. A combination of experiment and survey methods was used as instruments for data collection and is outlined in section 4.1.2.

4.1.1 Background studies

Analysis of cognitive and human factors in the DF process – building on knowledge from forensic science

As part of the literature review, relevant standards, best practice guidelines, and process descriptions were reviewed. They indicated a primary focus on technical aspects and physical handling of the exhibits, whilst human and cognitive factors and decision-making were absent or quite superficially mentioned. These insights, combined with the substantial body of research concerning the implication of the human factor from other forensic science disciplines, led to the research question of **Article 1**, which aimed to examine the risk of cognitive and human factors during the various steps of the DF process. The paper justified a need for more research on biasability and reliability issues in DF and suggested that the HEP framework (Dror, 2016) could be applied to examine and measure DF practitioners' reliability and susceptibility to bias during DF work.

Collecting submission forms

Based on my own experience and findings from my master's thesis, it was apparent that DF practitioners would often collaborate closely with investigation teams and that there was a culture characterised by information sharing rather than strict information management. Still, the experience was limited to a Norwegian law enforcement context. To ensure that the research was based on valid premises, submission forms and general information about commissioning procedures were collected from 30 units/organisations in Europe and the USA during 2018-2019 (Appendix 4). The objective was to understand how DF examinations were initiated, particularly concerning the task descriptions and task-relevant/task-irrelevant contextual information dissemination. An analysis of the collected material showed that it was common to convey contextual information either in the submission form or in a dialogue between the commissioning party and the DF practitioner. There were no delimitations or warnings against forwarding task-irrelevant information in any collected forms or procedures, which suggested that task-irrelevant information would sometimes be available to the DF practitioner before or during DF examinations.

4.1.2 The experiment

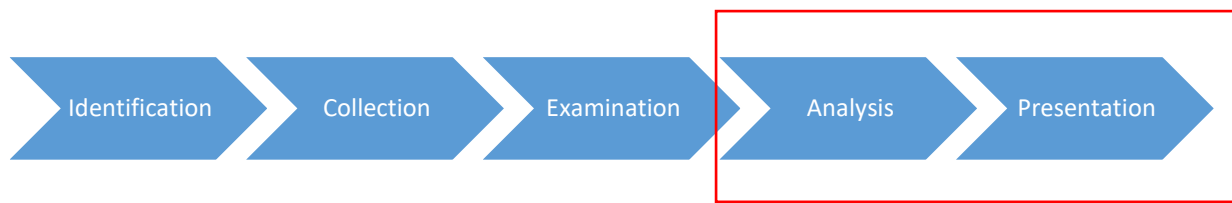
As the research question suggests, the aim is to explore the relationship between human factors (DF practitioner), the process (DF process), and the product (digital evidence). Different research traditions have developed methods to explore causal relationships. Positivism/empiricism-inspired traditions highlight that a researcher can most convincingly identify cause-and-effect relationships through experimental designs (Leedy & Ormrod, 2014, p. 234), and such a design was therefore chosen for exploring biasability in DF decision-making. Nevertheless, this does not rule out the fact that there are other methods for arriving at generalised causal relationships, such as intensive qualitative case studies or statistical manipulation and measurement (Shadish et al., 2002, pp. 392, 500).

Biasability and reliability are statistical constructs that cannot be directly observed or measured unless multiple actors are involved and compared. A quasi-experimental design (further referred to as *the DF experiment*) was developed to examine four distinct levels of the HEP framework (Dror, 2016). HEP 4 and HEP 8 are concerned with biasability between experts at observation and conclusion levels, and HEP 2 and HEP 6 are about reliability between experts at observation and conclusion levels. DF work is largely engaged with event reconstruction (see, e.g., Carrier & Spafford, 2004), and DF practitioners make inferences concerning what the traces mean individually or as inferences drawn, based on several traces

in context. These inferences would be more than factual descriptions about traces but, at the same time, not conclusions. This gap led us to define an additional level: *interpretation*, through which we explored the issues of biasability and reliability.

The aim of the DF experiment was twofold: first, to examine whether contextual information would bias DF practitioners' observations, interpretations, and conclusions and, second, to investigate whether DF practitioners who received similar conditioning information would achieve consistent results concerning their observations, interpretations, and conclusions. The DF experiment was conducted from October 2019 to January 2020 and centred on the analysis and presentation phases of the DF process (see Figure 1).

Figure 1: The DF process, adopted from Flaglien (2018).



a) Quasi-experimental design vs randomised experiment

The purpose of experiments is to test descriptive causal hypotheses about manipulative causes (Shadish et al., 2002, p. 14). The causal relationships examined in the DF experiment were the associations between the context and the observations, interpretations, and conclusions. The unique strength of experimentation is the ability to provide “causal description”, which entails “describing the consequences attributable to deliberately varying a treatment” (Shadish et al., 2002, p. 9). Yet, an experiment has less ability to clarify “the mechanisms through which and the conditions under which that causal relationship holds” (Shadish et al., 2002, p. 9), namely, the “causal explanations” that relate to *why* the consequences occur.

The common attribute in experiments is control of treatment (Shadish et al., 2002, p. 12). In a randomised experiment, participants are randomly assigned to groups that, on average, are probabilistically similar to each other. When given similar treatment, the observed outcome differences are likely to be caused by the treatment rather than differences between the groups that existed before the start of the study (Shadish et al., 200, p. 13). The term “quasi-experiment” refers primarily to the lack of random assignment and rests on the assumption

that the cause is manipulable and occurs before the effect is measured. Such experiments usually create less compelling support for counterfactual inferences, which means there could be many alternative explanations for the observed effect. However, several measures may be applied to rule out alternative explanations, in order to strengthen the validity of the estimate of the treatment effect (Shadish et al., 2002, pp. 13-14).

The DF experiment had a quasi-experimental design. It aimed to achieve high ecological validity by mimicking real-world working conditions as well as possible, while at the same time eliminating or controlling as many potential confounding variables as possible through a background survey and matching procedures, which are described and discussed below.

b) Choosing an evidence file

It was necessary to consider whether to use an actual or mock evidence file for the research. A significant challenge with a dataset obtained from an actual criminal investigation is that the ground truth is unknown, and it would be necessary to develop a gold standard. In addition, using such an evidence file appeared to be an ethical Gordian knot, which would require consent from the party from which it was obtained and from all third parties from whom content or communication was present. These ethical problems do not exist in a mock evidence file. Also, a well-documented mock evidence file includes oversight regarding which traces are present, and an investigation and validation of traces are not necessary to the same extent. The challenges of a mock evidence file relate to whether it has the “look and feel” of an evidence file obtained from an actual criminal investigation and that is not previously known to the participants, for example through training or competitions.

After reviewing several available evidence files, the evidence file from the “M57” scenario from Digital Corpora (S. Garfinkel et al., 2009) was selected. Although it was more than ten years old, it had several advantageous factors. First, the M57 case was not particularly grave and concerned an alleged information leakage of sensitive employee information. It was thus a case type that most of the participants could be assigned to in real life, as opposed to particularly grave or complex crimes that would probably be assigned to senior DF experts. Second, the evidence file was relatively small (3 GB) and would not require a lengthy processing time for the participants. Third, the evidence file was well documented, with an available “teachers’ guide”, in which the key findings were described in detail. Fourth, the material was not accessible online, and it was thus expected that the evidence file would be unknown to the participants. Fifth, the available traces and information did not point towards a single explanation, and it would not be possible to refute any of the relevant (offence level)

investigative hypotheses based only on the information on the evidence file. To remove any confusion around the old timestamps, the participants were asked to imagine that they were investigating the incident around the time indicated by the timestamps and not today.

c) Designing the scenario and conditions

The background study of submission forms and commissioning procedures (see section 4.1.1) justified that the DF practitioner would usually receive case information together with the task description. A basic scenario of reported information leakage from the company M57.biz was developed, in which the role of Jean Jones, the chief financial officer (CFO), needed further attention. The scenario included an illustration of the leaked document.

The scenario was as follows:

Confidential information leakage

M57.biz is a small US based company, with office in your country. The company, which develops and sells body art equipment (tattoo, piercing etc.), is in the start-up phase. The manager for the M57.biz office in your country is Alison Smith, and the CFO is Jean Jones. The company has 4 programmers, 2 in marketing, and 1 in business development. Only Alison and Jean have a permanent office space, while the other employees work from home office. All employees participate in a daily online meeting. There are in-person meetings for all employees in the M57.biz office once every two weeks. Most documents are exchanged by email.

A spreadsheet (m57plan.xls) containing confidential information was recently posted as an attachment in a forum of a competitor’s website. When this was discovered, Alison reported the incident to the police as information theft. Alison told police that Jean, the CFO, was responsible for updating the spreadsheet, and that it was probably sent from Jean’s computer.

The attachment posted on the competitor’s website looked like this:

M57.biz company				
Name		Position	Salary	SSN (for background check)
Alison	Smith	President	\$140,000	103-44-3134
Jean	Jones	CFO	\$120,000	432-34-6432
Programmers:				
Bob	Blackman	Apps 1	90,000	493-46-3329
Carol	Canfred	Apps 2	110,000	894-33-4560
Dave	Daubert	Q&A	67,000	331-95-1020
Emmy	Arlington	Entry Level	57,000	404-98-4079
Marketing				
Gina	Tangers	Creative 1	80,000	980-97-3311
Harris	Jenkins	G & C	105,000	887-33-5532
BizDev				
Indy	Counterching	Outreach	240,000	123-45-6789
Annual Salaries			\$1,009,000	
Benefits			30%	\$302,700

Whilst the former research has led to a recommendation not to provide task-irrelevant information to the forensic analysts (Forensic Science Regulator, 2020), determining what is task-relevant in DF casework is not straightforward. When designing the DF experiment, I could not identify studies discussing or informing this topic within the DF domain, nor any guidelines or standards targeting the DF discipline clarifying the issue. A study targeting the general forensic science domain showed high variation in what forensic scientists considered task-relevant information (Gardner et al., 2019). Due to the lack of clear guidance on what should be considered a task-irrelevant context, the contextual information developed for the experiment was assumed to be task-irrelevant or at least not task-relevant to the specific case they were assigned to examine.

In addition to the scenario described above, additional *contextual information* indicating guilt (strong), guilt (weak/ambiguous) and innocence was provided to the experimental groups, except the control group, who received no additional context:

- **Control** – no information
- **Guilt condition** – *Jean was arrested for information theft, and in a police interview earlier today, she confessed that she had committed the criminal offense.*
- **Weak guilt condition** – *Alison told police that there has been a long-lasting wage dispute with the programmers in the firm, who claim to be underpaid. One of the programmers – Emmy Tuckford Arlington – has fronted the dispute on behalf of the programmers in M57.biz. Jean has supported the programmers in this conflict, and has told Alison that the company can afford to pay them better salaries. Jean is about to be interviewed by the police about the reported incident. However, the chief investigating officer wants an analysis of Jean's computer before the police interview, to look for traces indicating that she was involved in the reported incident.*
- **Innocence condition** – *As a result, Jean was arrested for information theft and questioned about the incident in a police interview. However, after the police interview, the police believe she is innocent, and that she was framed during a phishing attack.*

d) Recruitment/sample

The participants were recruited through various approaches. For the Norwegian participants, the managers at the respective DF units were contacted and asked for the contact information of potential participants fitting the *profile description*: “The project is mainly aimed at digital

forensic examiners who have digital forensic work as their primary task.”¹ When they returned a list of potential participants, all were contacted through a personal email, with an invitation to participate in the research. An information letter with a consent statement (Appendix 5), and the survey of background variables (Part 1 survey) were enclosed with the invitation. The DF practitioners were encouraged to return the signed consent form and complete the Part 1 survey if they agreed to participate. A reminder was sent to those invited DF practitioners who had not responded within a week. To recruit participants from other countries, invitations were sent through professional networks such as the Europol-hosted European Union Cybercrime Task Force (EUCTF) forum, the INTERPOL Digital Forensic Expert Group, the European Cybercrime Training and Education group (ECTEG), and the national cybercrime units in Europe, Australia, and the USA. Invitations were also forwarded through research fellows and professional contacts. This massive recruitment campaign resulted in many responses regarding interest and 65 consenting participants, the majority from Norway. A total of 56 DF practitioners completed the experiment. Three of these were excluded, since they did not fit the DF practitioner profile, which resulted in 53 DF practitioners from eight countries represented in the experiment sample: Norway (44), UK (2), India (2), Canada (1), Denmark (1), Finland (1), Kenya (1) and The Netherlands (1).

e) Survey of background variables (Part 1 survey)

When signing up for participation, the DF practitioners were asked to complete a survey (Part 1 survey – see Appendix 6) with eight questions concerning their gender, professional background, educational qualification/degree, organisational level, professional experience, and preferred analysis tools. The purpose of this survey was twofold. First, it was used to match participants in experimental groups, to ensure that the distribution in the respective groups was as similar as possible (see 4.3.2 f, for procedure details) and, second, to gain insight into the characteristics of DF practitioners, based on the included variables.

f) Matching procedure

As described, random assignment to groups is the preferred procedure when conducting experiments, which (given a sufficient number of participants) enables the researcher to reasonably assume that the groups, on average, are similar and that the differences are entirely due to chance (Leedy & Ormrod, 2014, p. 238). When other variables may influence the dependent variable, *matching* is a measure to ensure similar groups with respect to such

¹ Prosjektet retter seg i hovedsak mot dataetterforskere som har datateknisk etterforskning som sin primær oppgave (for Norwegian participants)

variables before the experiment is carried out (Leedy & Ormrod, 2014, p. 238). It was assumed that factors such as educational background (police vs civil), educational level (PhD, MSc, BSc, other) organisational level (local, national) and years of experience with DF investigation could influence the results. It was also uncertain how many participants we would be able to recruit. It is preferable to ensure an equal distribution before the experiment, instead of improving comparability statistically after the experiment is done (Shadish et al., 2002, p. 353). Based on the above-mentioned criteria and the responses to the background survey (Part 1 survey), the participants were matched in groups of four, to ensure equal distribution in terms of background. They were then randomly assigned to one of the four experiment groups based on a lottery, by drawing from a bowl bearing numbers from one to four.

g) Experiment conditions

As described, the DF experiment design was a compromise between mimicking real-world working conditions and eliminating or controlling confounding variables. In addition to the aforementioned matching procedure, there were numerous considerations and decisions related to the variables, such as time frame, date, place, software, and templates, which are discussed below.

In terms of time frame and date, the DF practitioners were asked to reserve half a day (4-5 hours) for the experiment. This recommendation was based on several considerations. Since the DF experiment was aimed to examine biasability and reliability, it was essential to ensure as similar conditions as possible and thus control the time variable. It was crucial to recruit enough participants to have adequately sized experimental groups and statistical power in the results. It was assumed that, if completing the experiment required much time and effort, it would be difficult to recruit sufficient participants, and there would be a high risk of attrition. On the other hand, the participants should have the necessary time to complete the analysis and write a report. To ensure that the participants could use the time effectively, they were allowed to download and process the evidence file the day before, which enabled them to focus solely on analysing and reporting on the experiment day. For comparison, the estimated time frame for completing the NIST black-box study was two hours per disk image (Guttman et al., 2022, pp. 38-39).

Another trade-off was whether the DF experiment should be arranged on a fixed or flexible date. A fixed date would prevent information about the experiment from reaching participants that had not already completed it. At the same time, it would pose a high risk of failure if

something went wrong and a high risk of attrition if they needed to prioritise an urgent request in real-life casework. All involved participants were busy professional DF practitioners, and a flexible date would probably reduce attrition due to the need to prioritise urgent casework.

Based on these considerations, the participants were allowed to choose the date to conduct the experiment. To minimise the risk of biasing participants who had signed up for a later date, they were instructed not to talk about the experiment and their findings.

To control the environmental variables, the DF experiment could have been arranged within a laboratory environment with similar hardware, software and control over interaction between participants. However, completing the experiment under such conditions was likely to differ largely from how the DF practitioners would work during an actual investigation.

Consequently, it would potentially increase the research effect and limit the ecological validity. Using unfamiliar analysis software would potentially introduce extra time constraints, due to the new graphical user interface and functionalities. Therefore, it was decided that the participants would conduct the DF experiment at their regular workplace, with their typical hardware and analysis software, which would be close to how they would usually work in a real investigation. On the administrative side, conducting the experiment in a lab would require software licence costs and travelling expenses for the participants, and the PhD project had no funding for covering such expenses. As a substitute for controlled lab conditions, the DF practitioners were instructed to work alone during the DF experiment. Information about which analysis software they had used was collected in the Part 2 survey provided to the DF participants right after the experiment was completed. This information enabled a statistical control for the analysis software variable at a later stage.

A blank template was provided for the analysis report, and they were encouraged to write the report in the same manner as in actual casework. They were given a log template for making notes during the analysis and were asked, as a minimum, to note when the analysis started and ended. Further, they were asked to bookmark findings and export them to a PDF file if their analysis software offered such functionality.

h) Pilot study

Two independent and experienced DF practitioners were involved in the pilot study. The first DF practitioner reviewed and validated the traces described in the teacher's guide and had no objections. The second DF practitioner volunteered as a test pilot in the experiment prior to rolling it out to the consenting participants. The review resulted in a few corrections to the documentation, to improve clarity. A potential challenge with the chosen platform was

discovered. Consequently, the evidence file was moved to a platform (Google Disc) that allowed the DF participants to access and download the evidence file without signing up for a service.

i) Preparation for and completing the DF experiment

After agreeing to participate and returning a signed consent form, the date for completing the experiment was settled, and the participants were informed about how to prepare for the experiment day. The link to the evidence file was sent to the participants the day before the experiment, and they were allowed to download and process the file in advance. They were reminded not to start the analysis until they had received the necessary information, instructions, and templates the next day.

On the experiment day, the DF practitioners received an email with the information, instructions, and templates. The Norwegian participants received documents in the Norwegian language, while all others received them in English:

- “READ THIS_ Description of the experiment” / “LES DETTE_ Beskrivelse av gjennomføringen” – a document with the scenario and task description (Appendix 7-10).
- “C_Log” / “C_Logg” – a word template for taking contemporaneous notes during the analysis (Appendix 11).
- “E_Report from analysis” / “E_Analyserapport” (Appendix 12) – a blank template for the analysis report.

The READ THIS_ Description of the experiment document described the scenario and contained one of the four versions of contextual information described in 4.3.2 c). All received the same task description: “*You are tasked with analysing a copy of the hard drive from Jean’s computer and find out: **What has happened, and what was Jean’s involvement in the reported incident?***” They were reminded to work alone and not consult or confer with anyone during the analysis. After completing the analysis, the DF practitioners returned the analysis report, bookmarked traces (optional) and log.

j) Post experiment survey (Part 2 survey)

Immediately after completing the DF experiment and submitting the documents, the DF practitioners received the Part 2 survey. Here they were asked whether they had past knowledge or experience of the scenario, to have the possibility to eliminate such participants from the sample. Then they were also asked to report which analysis software they had used.

The following questions concerned their analysis approach, namely:

- what they believed had happened, after reading the scenario (and conditioning information) and before starting the analysis,
- what they believed had happened, after completing the analysis,
- which techniques they used to maintain objectivity during the analysis and which techniques they used to control evidence reliability.

The next question concerned their conclusion about the findings. They were asked to consider 17 traces and indicate, for each, whether they found it or not and – if they did – whether the trace indicated guilt, innocence, or ambiguity regarding Jean, the suspect. Since they had already submitted their analysis report, they had no opportunity to change the report based on the information presented here. They could also annotate any unlisted relevant traces and rate them in terms of guilt, innocence, or ambiguity (see full description of Part 2 survey in Appendix 13). Finally, they were asked an open-ended question about comments or remarks.

k) Attrition

Even if an experiment starts with equal distribution in the groups, attrition can influence the result (Leedy & Ormrod, 2014, p. 237). Sixty-five DF practitioners consented to participate. Four withdrew before the agreed experiment day. Four started the analysis and withdrew before completion. Three were excluded from the sample because they did not fit the profile of a DF practitioner. Fifty-three completed the experiment, and all submitted a complete set of documents (Part 1 survey, analysis report, log, bookmarks (optional) and Part 2 survey). None withdrew after completion. Table 1 shows the attrition in the DF experiment per experiment group.

Table 1: Attrition in the DF experiment according to the received context.

	Control	Strong guilt	Innocence	Weak guilt	Total
Completed	16	12	12	13	53
Agreed to participate	17	15	15	18	65
Drop out before start	1	1	-	2	4
Drop out after start	-	1	1	3	5
Excluded from sample	-	1	2	-	3
Total attrition	1	3	3	5	12

l) *Summary of the collected material*

As a result of the background studies and DF experiment, the following material was collected (see Table 2).

Quantitative descriptions may be helpful to shed light on the magnitude and richness of the information obtained from the experiment. Since the participants each took 4-5 hours on the experiment, the documentation encompasses a total of 265 hours (not including the time for processing the evidence file) or approximately 33 full-time (eight-hour) workdays of DF investigative work at the analysis and presentation stages of the DF process.

Table 2: Overview of the collected material.

Material	Description	Relevance to the articles
Documents	Various normative national or international best practice guidelines and standards describing procedures for handling digital information in the context of a criminal investigation.	All
Documents	Submission forms (N=30) and/or commissioning procedures (N=11) obtained from law enforcement agencies or international organisations in Europe and the US.	2
Documents	Analysis reports (N=53), logs and bookmarks produced by DF practitioners during the DF experiment. The logs and bookmarks were not subject to further analysis.	2, 3 and 5
Survey	Background survey (Part 1 survey) on variables related to professional background, education level, experience, and user experience with analysis software (N=53).	2, 3, 4 and 5
Survey	Post experiment survey (Part 2 survey), with a combination of closed and open-ended questions about how they approached the analysis of the evidence file on how they assessed the findings in terms of guilt, innocence, or ambiguity (N=53).	2, 4

The analysis reports constitute a total of 248 pages, ranging from 1-12 pages per report, with an average of 4.7 pages per participant. The Part 2 survey was also a rich and detailed source of information about investigation practice. The focused questions resulted in targeted responses. The material made up a total of 11 pages of information of the participants' written accounts of their own approach towards handling context, safeguarding examiner objectivity and evidence reliability, which on average was approximately 1/5 page per participant.

4.1.3 Strengths and limitations of the research design

Although the characteristics of the total DF practitioner population are unknown, the DF experiment sample and material are considered adequate and representative of DF practice. The participants were all DF practitioners, with a variation in gender, educational background, competence, organisation level, and experience with DF work. Eight countries were represented in the study, but, due to the high proportion of Norwegian participants (83%), it is not possible to rule out a bias resulting from investigative procedures, practices, or mindsets, which may be particular to Norwegian DF practitioners. Yet, no factors pointing in such a direction were observed. The DF experiment had high ecological validity, due to the experimental design mimicking actual DF casework: They received some background information about the case and a task description, which, according to the survey of submission forms and commissioning procedures, is how the actual casework would also be initiated. They processed and analysed the evidence file with tools and methods they would typically use and were encouraged to use their standard structure or template for the report.

The results presented in **Article 2** indicate biased observations of traces associated with the received context. The sample size is relatively small for statistical analysis, limiting the statistical power in the mere quantitative analysis of biasability in **Article 2**. The rule of thumb is 20-24 participants per condition (Brysbaert, 2019 p. 19), and a minimum of 12 per condition was decided. Since this is the first experiment aimed at testing bias in DF decision-making, more studies are necessary for a more solid underpinning of the relationship between contextual information and biased observations. However, the small sample size does not influence the reliability calculation between practitioners receiving a similar context. The Krippendorff Alpha Coefficient calculations are flexible and allow for small or large sample sizes (Hayes & Krippendorff, 2007; Krippendorff, 2011).

The time condition was considered in the planning and was decided as a compromise between the possibility to recruit busy DF practitioners and the risk of high attrition on one hand and the risk of too little time to complete the analysis and reporting on the other. Only four participants mentioned time constraints in Q9 of the Part 2 survey, but 13 of the reports lacked a final conclusion or seemed unfinished in terms of language and style.

If the participants had been given no time limit, more would probably have observed a higher number of traces. Still, as indicated in the Part 2 survey (Q7), there were many more potentially relevant traces than those used for benchmark comparison between the groups. An unlimited time frame would thus have been beneficial to all participants and not only those

who found the least number of traces. Only traces discovered by at least 31% of a group were included in the test for biased observations, interpretations, and conclusions. Time constraints could have led to poorly articulated descriptions or conclusions concerning the evidence but would still not have influenced whether they were coded as observations. For example, as long as the report mentioned the file m57biz.xls, it was coded as observed, regardless of how well or in how much detail the file was described or visualised.

The time limitation may have influenced the number or scope of techniques for examining or controlling evidence reliability referred to in the Part 2 survey and **Article 4**, as well as the documentation of these in analysis reports explored in **Article 5**. Although correspondence between quality control and documentation practice is expected, this is merely an assumption. To account for unfinished reports due to time limitations, only reports that included a summary or final conclusion were included in the sample for **Article 5**. Reports that seemed to be, or stated that they were, unfinished or lacked a final conclusion, were excluded for this particular purpose.

The Part 2 survey obtained the participants' own accounts of how they approached the analysis of the evidence file, and which techniques they used to safeguard examiner objectivity and evidence reliability. Their responses depended on what they perceived to be a *technique* and what they remembered having applied, but could also have been influenced by what they knew they *ought to* apply to perform the task properly.

The evidence file was manufactured and did thus not originate from an actual investigation. Much effort was invested in choosing an adequate evidence file that would include what one would expect to find in actual casework. Still, the participants knew that what they did or did not discover would not lead to any consequences for a suspect. Further, the reports would not undergo scrutiny in a court hearing, which might have led to less effort or thoroughness during the analysis and documentation phase. On the other hand, knowing that they were participating in research and that a researcher would read their reports might have led to a higher focus on themselves, i.e., the quality of their own analysis and documentation process, rather than on the evidence, compared to actual casework.

The participants were invited to participate, which might have led to systematic deviance in the characteristics of DF practitioners that accepted the invitation and those who declined, due to a self-selection bias. Many DF practitioners did not reply to the email, and, of those who

replied, many did not provide a reason for declining. Those who agreed to participate were not asked why. There is, thus, an insufficient empirical foundation for drawing assumptions about significant differences between these groups.

4.2 Analytical procedures

The analysis for **Article 2** was planned before the DF experiment was carried out. The ideas for **Articles 3** and **4** emerged from an exploration of the rich empirical material collected through the DF experiment and related surveys. The idea for **Article 5** came about after discovering the research of Bali et al. (2020, 2021), which provided an opportunity to examine issues such as opinion types and report content in DF practice relative to other forensic science disciplines. An overview of the material, type of analysis, and which article they relate to is presented in Table 3.

Table 3: Overview of the material, type of applied analysis, and associated articles.

Material	N=	Type of analysis	Article	Comment
Submission forms and commissioning procedures	30/11	Quan/qual	2	
Analysis reports (all)	53	Quan	2	
Analysis reports (all)	53	Qual	3	
Analysis reports (sample)	40	Quan/qual	5	Only reports that included a final conclusion/summary
Part 1 survey	53	Quan	2	Also referred to in 3, 4 and 5 – but not used for measurements
Part 2 survey (question 1, 2, 7, 8)	53	Quan	2	
Part 2 survey (question 3, 5, 6, 9)	53	Quan/qual	4	

The thesis's articles form the structure of this section. First, the analysis for **Article 2** is presented, which included submission forms and commissioning procedures obtained prior to the DF experiment as well as analysis reports and survey results collected during the DF experiment. Then, the qualitative analysis for **Article 3** is presented, followed by the combined quantitative and qualitative analyses for **Articles 4** and **5**. As **Article 1** was subject to a theoretical analysis and discussion, and not the result of data collection, it is not elaborated further here.

4.2.1 Article 2

Background survey – submission forms

Submission forms and commissioning procedures were collected to learn whether a DF practitioner would usually receive contextual information about the case and whether any restrictions were conveyed concerning task-relevant or task-irrelevant contextual information, either in writing or through dialogue. The analysis was performed in MS Excel, and annotations were made regarding whether the forms provided for the sharing of case information, whether there were any restrictions on what to share/not to share, and whether dialogue about the task was used instead of or in addition to the form. Quantitative measurements were used for calculating proportions in percentages, and the results were reported in **Article 2** pp. 2-3.

Statistical analysis of analysis reports and survey responses

The 53 analysis reports and responses to Part 1 and Part 2 surveys collected during the DF experiment were analysed with the objective of informing the following research questions:

1. *Are DF examiners biased by contextual information when making observations, interpretations of observations, or in their conclusions during the analysis of digital traces?*
2. *Are DF examiners consistent with one another when making observations, interpretations of observations, or conclusions during the analysis of digital traces?*

The material was coded in IBM SPSS Statistics (Release 26.0.0.0. 64-bit edition), and the dataset was published together with the paper. Descriptive statistics were conducted for the demographics of the sample. In the Part 1 survey, background variables were coded in SPSS. The traces found by at least 31% were included in the statistical analysis of biasability or reliability of observations, interpretations, or conclusions.

The analysis reports were used to identify and code *observations of traces*. Eleven individual traces were selected, with the values “identified” or “not identified”. To be coded as “identified”, the trace had to be mentioned in the analysis report. This criterion entailed that seeing the trace was not enough; the participant needed to consider it relevant and include it in the analysis report. Nevertheless, there is a discrepancy between the proportion of observed traces (section 3.1.1 in **Article 2**) and the proportion of those that stated having identified the trace in the Part 2 survey (Appendix 3 in **Article 2**), which was completed after submitting the

analysis report. A plausible reason for the divergence is that, although several participants did not consider it relevant during the analysis, they might still recall having seen the trace when confronted with it in the survey.

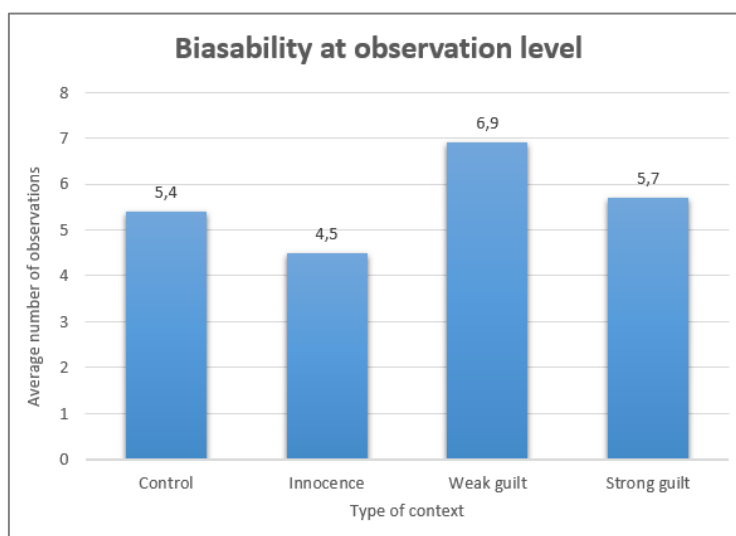
Seven essential *interpretations of the observed traces* were used for the statistical analysis. Similar to observations, the analysis reports were used to identify and code interpretations of observed traces, with the values “identified” or “not identified”.

In the Part 2 survey, responses to questions 2, 7, and 8 were included in the analysis. In question 2, they were asked to convey which analysis program(s) they had used. The purpose was to control for a possible biasing effect of the particular software, but no such effect was found. Question 7 concerned their *conclusions*. For each of the 17 listed traces, they were asked to tick off whether they found the trace and, if so, how they assessed it in terms of guilt, innocence, or ambiguity. In question 8, they could add additional relevant traces to the list and rate them.

The participants were asked to export bookmarked findings into pdf documents and use the log template. However, these documents were excluded from the analysis because not all analysis software facilitates bookmarking, and log usage was very inconsistent. It could be argued that these documents held accounts of observations, but, similar to observations reported in the Part 2 survey, they were not considered to fulfil the criteria of “observed and deemed relevant to report”.

The choice of *statistical test* hinges on the correspondence between the data and the test assumptions. The Kruskal-Wallis test, which is the non-parametric alternative to the One Way ANOVA, was chosen to measure differences between the experimental groups. Non-parametric statistics are appropriate for data on the ordinal level, can be used for skewed populations, and are suitable for relatively simple analyses (Leedy & Ormrod, 2014, p. 294). The results indicated contextual bias in the observation of traces but no significant association between the context and interpretation of observed traces and conclusions (see Figure 2).

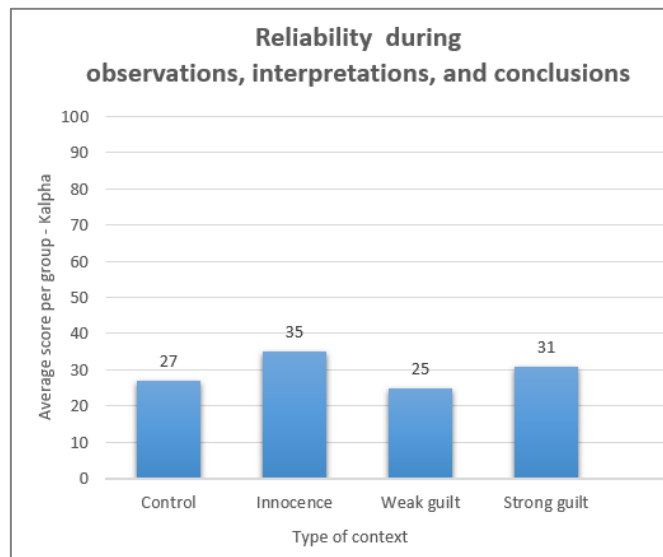
Figure 2: Average observed traces (of max. 11) per group in **Article 2**.



To test reliability, the participants who received the same information were compared. There are several tests to choose from, such as percent agreement, Bennett et al.'s S, Scott's pi, Fleiss's *K*, Cohen's kappa, and Krippendorff's alpha (Nili et al., 2017). Krippendorff's alpha (α) was considered the best fit, due to its flexibility for comparing multiple coders, measurement levels, and sample sizes, as well as its ability to tackle incomplete or missing data (Hayes & Krippendorff, 2007; Krippendorff, 2011; Nili et al., 2017). The tests showed, as indicated in Figure 3, low reliability for all groups at all measured levels, namely, observations of traces, interpretations of observed traces, and conclusions.

The observations of traces and interpretations of observed traces were based on participants' descriptions in their analysis reports. In contrast to coding the surveys, this coding required some interpretation. An independent coder coded 10% of the reports, to control coding agreement. Krippendorff's alpha test was used to compute the inter-coder reliability, with the result $\alpha=0.91$, which is considered a strong level of agreement.

Figure 3: Average reliability score per group for the observation, interpretation, and conclusion levels in **Article 2**.



4.2.2 Article 3

The following was the research objective of Article 3:

The aim is to explore whether and how the evidential value is crafted by the DF practitioner and to shed light on how the evidence elasticity enables the DF practitioner to turn the traces into misinformation with the propensity to mislead the legal decision-maker's assessment of evidential value.

The analysis centred on the report descriptions' diversity. A thematic analysis (Braun & Clarke, 2006) of the reports was performed in QSR International NVivo Pro Edition, Version 12.1.1.256 64 bit. Some themes were already triggered during the review of reports and coding for the statistical analysis of the DF experiment, which led to a deductive approach to the analysis of the reports. The inductive analysis focused on additional variance elements, not directly related to the statistical measurements. A list of themes emerged from the analysis, as shown in Table 4.

Table 4: Overview of themes and codes in **Article 3.**

Themes	Codes
Interpretations of the assignment	<ul style="list-style-type: none"> - Assignment consistent with the original - Assignment differing from the original - Missing account of the assignment
Reconstructing what had happened	<ul style="list-style-type: none"> - Search for and/or observations of the leaked spreadsheet (m57plan.xls) on the suspect’s computer - Observing the file m57biz.xls on the suspect’s computer - Comparing the leaked file (m57plan.xls) with the file m57biz.xls found on the suspect’s computer
Determining who were involved	<ul style="list-style-type: none"> - Activity description related to entities - Activity descriptions related to persons
Constructing conclusions	<ul style="list-style-type: none"> - What? <ul style="list-style-type: none"> o CFO fraud o Phishing/spear phishing o Hacking o There was no information leakage - Who? <ul style="list-style-type: none"> o Jean had caused/contributed to the information leakage o An insider caused the information leakage o An outsider performed the information leakage - How? <ul style="list-style-type: none"> o The info was sent by email o The info was copied out on a USB o The info was obtained through hacking o No information leakage had happened

4.2.3 Article 4

The DF practitioners responded to the “Part 2 survey” immediately after completing the experiment and submitting their analysis reports. The responses were subject to a combined quantitative–qualitative analysis. The 53 survey responses to four open-ended questions provided the participants’ *own accounts* concerning how they approached the analysis after receiving the scenario information (and biasing context) and whether and how they applied techniques to safeguard examiner objectivity and control evidence reliability during the analysis. The fourth question allowed them to provide any comments or remarks. This material was a valuable supplement to the other material obtained during the DF experiment, since it largely resulted from investigative activities they had just performed.

The material was analysed to examine the following research question:

How do DF practitioners handle contextual information, approach examiner objectivity and evidence reliability during the analysis of an evidence file?

When exploring the responses to the question concerning the scenario/contextual information, it became apparent that most of the participants mentioned scenarios or hypotheses. A quantitative approach (performed in MS Excel Office Professional Plus 2016) was applied, to examine the proportion that used hypotheses, the number of hypotheses they had developed, and how many included an innocence hypothesis.

When analysing the responses to the other questions concerning techniques for examiner objectivity and evidence reliability, the responses were first reviewed for whether they mentioned a technique or not. The responses that mentioned techniques were coded for which techniques they mentioned, and similar or related techniques were assembled in the same code group. The analysis resulted in eight code groups for examiner objectivity techniques and seven code groups for evidence reliability techniques; see Table 5.

Table 5: Code groups of techniques and approaches in **Article 4**.

Activity	Code groups of approaches and techniques
Handling of contextual information (Q3)	<ul style="list-style-type: none"> - Used hypotheses (yes/no) - Number of hypotheses (number) - Included innocence hypothesis (yes/no)
Examiner objectivity techniques (Q5)	<ul style="list-style-type: none"> - None - Hypotheses - Focus on facts - Avoidance - Information requirement - Forensic procedures - Neutral presentation - Other
Evidence reliability techniques (Q6)	<ul style="list-style-type: none"> - None - Dual tool verification - Metadata examination - Cross-check of findings - Hash calculation - Manual verification of output - Timeline analysis

4.2.4 Article 5

Article 5 results from inductive discoveries during the quantitative analysis for **Article 2**. It became apparent that there were significant variations in how DF practitioners articulated their conclusions and how they documented information about processes, tools, and procedures in their analysis reports. Another factor that motivated this paper was a published study of reporting practices in eight other forensic science disciplines (Bali et al., 2020, 2021). In addition to merely describing the characteristics of the DF discipline, the work of Bali and colleagues provided an opportunity to explore DF practice in contrast to other relevant disciplines and identify similarities and differences. Bali et al. studied a sample of 500 reports sourced from forensic proficiency tests from eight different forensic science disciplines, with a median number of reports from the included disciplines of 56, ranging from 36 to 121. The 40 DF reports were considered a suitable quantity for comparison.

The following research questions guided the analysis:

What characterises the opinions used in DF reports in terms of opinion type, uncertainty expressions, addressed issues, and included content relevant to the credibility assessment of the reported results? To what degree does the DF reporting practice concerning applied opinion types and included content deviate from other FS (forensic science) disciplines?

The analysis procedure for exploring DF reporting practices was threefold:

- a) a quantitative analysis of conclusion types and level of issue (source, activity, offence)
- b) a quantitative analysis of report content
- c) a qualitative analysis of (un)certainly descriptions

Analysis of conclusion types and level of issue

Bali et al. (2020, 2021) performed a quantitative content analysis of conclusions and content relevant to the result's credibility. The coding from this study was applied and expanded with additional coding categories for the study of DF reports. The conclusions were coded in IBM SPSS Statistics (Release 26.0.0.0. 64-bit edition) according to six conclusion types, see Table 6.

Table 6: Conclusion types and values used in the quantitative analysis in **Article 5**.

Conclusion type	Values
Categorical conclusions	Only traditional, only elaborated, both
Likelihood ratios (LR)	Numerical, verbal, both
Random match probabilities (RMP)	Observed, not observed
Likelihood of observed similarity statements (LoS)	Observed, not observed
Strength of support statements (SoS)	Observed, not observed
Source probabilities (SP)	Numerical, verbal, both

The conclusions studied in Bali et al.'s (2020) research were focused on feature comparison procedures and would thus primarily address source level issues. To gain more insight into the issues targeted by DF investigations, the conclusions were analysed for whether they addressed issues at source/sub-source level, activity level, or offence level (values: only crime category, only guilt/intention, both). Since it sometimes was difficult to differentiate between what was a justification of the conclusion or the conclusion itself, all the narrative under the heading "Conclusion" (or other variations such as "Result", "Main conclusion", etc.) was coded for the addressed issue level, as long as it related to aspects concerning the relevance or credibility of the information. This approach implied that several issue levels could be observed within the narrative of a single conclusion.

Analysis of report content

A quantitative content analysis was performed based on 13 categories of content. Eight categories were obtained from Bali et al. (2020) (marked with *), and five additional content categories were applied due to their relevance to the DF discipline, see Table 7.

Table 7: Content type and values used in the quantitative content analysis in **Article 5**.

Content type	Values
Description of the analysis methods used*	None, vague, specific
Information about the method/tool reliability*	Observed, not observed
Information about the method/tool validity*	Observed, not observed
Limitations of the methods or conclusions*	Observed, not observed
Reasoning or justification for the final conclusion (e.g., information about frequency or similarities)*	Observed, not observed
Any alternative explanations for the results*	Observed, not observed
Additional explanation of jargon or scientific terminology*	Observed, not observed
Explication that the conclusion is an opinion*	Observed, not observed
Task description	None, inaccurate/vague, correct
Description of received contextual information	None, inaccurate/vague, correct
Description of analysis tools used	None, inaccurate/vague, specific
Qualification or competence	None, inaccurate/vague, specific
Time zone	None, only abbreviation/name, abbreviation/name and explanation of how it should be understood compared to local time

Analysis of (un)certainty expressions

A qualitative analysis of the conclusions was performed to explore the usage of expressions describing (un)certainty. The analysis aimed to explore diversity rather than frequency. All the narrative within the conclusions was included in this analysis, as long as it related to aspects concerning the relevance or credibility of the information that was addressed. All the phrases from the conclusions describing (un)certainty within the above-mentioned criteria were extracted and sorted, and duplicates were excluded.

4.3 Research quality and ethical considerations

4.3.1 Permissions

The application to the Norwegian Police Directorate was approved on 08.08.2019 (Appendix 1). The application to the Norwegian Attorney General was approved on 15.01.2020 (Appendix 2). The latter concerns the collection of reports from actual cases and interviews with report authors, which have been postponed to a prospective post-doc project. The

research project was reported to the Norwegian centre for research data (NSD) (application number 458568) and approved on 12.09.19 (Appendix 3).

4.3.2 Anonymisation and handling of the material

The collected submission forms did not contain any case information, and did therefore not require any confidentiality protection. To ensure availability, a backup of the material was stored in a separate folder.

In terms of the material from the DF experiment, the participants were assigned a participant number. Any participant names were deleted from the analysis reports, and the collected documents were named according to the participant's number. The form associating names and participant numbers was stored separately from the anonymised survey responses and reports. A backup of the material was stored on an external hard disk.

4.3.3 Research quality

Validity (internal and external) and reliability are central aspects to consider for the parts of the research involving counting and calculations. The measures applied to achieve internal validity and inter-coder reliability were discussed in the respective articles and sections 4.1.2, 4.1.3 and will thus not be repeated. Still, some overarching aspects remain to be discussed. These are the value of pre-registration of research, triangulation, external validity and ethical considerations.

Pre-registration

Pre-registration is a measure for strengthening the research quality and involves registering predictions, research plans, and analysis plans on an online platform *before* the results are known (see, e.g., Chin et al., 2019, pp. 269-274; Searston et al., 2019). Once the pre-registration is submitted to an online platform, it is timestamped and uneditable (Chin et al., 2019, p. 269). This measure may reduce duplicative research efforts, protect the statistical integrity of studies by reducing flexibility in analysis and “p-hacking”, and improve access to research findings (Kimmelman, 2021, p. 645). Pre-registration is required in some medical research areas and becoming increasingly popular in other fields, including forensic science (Chin et al., 2020; Chin et al., 2019, pp. 269-272). Unfortunately, the DF experiment was not pre-registered, since I was unaware of this measure when designing the experiment. However, the dataset was published together with **Article 2** for transparency concerning the variables and values used for the statistical calculations.

Triangulation

Triangulation involves applying two or more methods to investigate the same phenomenon and answer the same central research question (Heap & Waters, 2019, p. 111). The objective of triangulation depends on the ontological position. From the realist (positivism/empiricism) position, triangulation can be used to check the validity of each component. This approach is referred to as “basic triangulation” and rests on the assumption that, when data from different components are used, invalid inferences due to errors and bias within the data are less likely (Heap & Waters, 2019, p. 112). Basic triangulation was essential for the internal validity of **Article 5**. The results from the DF experiment were compared with a dataset collected by Bali et al. (2020) involving practitioners from several forensic science disciplines. During the publication process of **Article 5**, I discovered possible errors in the calculations in the external dataset and contacted the corresponding author. The calculations were corrected, and a corrigendum (Bali et al., 2021) was published before **Article 5** was published. This showed that errors, which in this case were a practitioner’s unintended mistake (see section 3.4.1), can happen despite the scrutiny of scientific journals’ peer-review processes.

From a relativist (constructivism/interpretivism) perspective, “indefinite triangulation” may be used to obtain various accounts of the same phenomenon or event, by applying multiple methods to the same question (Heap & Waters, 2019, p. 112). The mixed-methods design applied in the thesis may be considered indefinite triangulation, since both qualitative and quantitative methods were used to collect empirical material (experiment and qualitative/quantitative survey) and analyse it from different perspectives (statistical/quantitative, qualitative), with the objective of exploring the observed phenomena further, rather than controlling their validity. Indefinite triangulation was a vital measure for informing the research question. For example, by exploring variability (low reliability), discovered through the statistical analysis in **Article 2** and further in quantitative and qualitative analyses in **Articles 3** and **5**, it was possible to theorise not only about the fact that low reliability was found but also about the range of the variance and how it related to the components forming the evidential value.

In mixed-methods research, triangulation may also be used for “epistemological dialogue” where different methods not only provide different kinds of information about the same phenomenon but constitute the world in different ways (Heap & Waters, 2019, p. 112), and this approach was applied in the thesis. On one hand, the statistical results in **Article 2** suggested low reliability between DF practitioners that were analysed under similar

conditions. Intuitively, this may be assumed problematic for the DF discipline, since an analysis would potentially lead to different results, depending on which DF practitioner was assigned to the task. Yet, exploring the variability from different epistemological perspectives enabled a more nuanced view of the variability, namely, that it may also be beneficial to the investigation – which is discussed in section 6.3.4.

External validity

Validity refers to an inference's approximate or tentative truth and is thus not a property of methods and design (Shadish et al., 2002, p. 34). External validity is often referred to as generalisability and is concerned with “the validity of inferences about whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables” (Shadish et al., 2002, p. 38). The question is thus whether the research involving the 53 DF practitioners has value outside this group.

The targets of generalisation may be from narrow to broad, broad to narrow, at a similar level, to a similar kind, to a different kind and from a random sample to a population (Shadish et al., 2002, pp. 83-84). The DF practitioners were not drawn from a random sample, which excludes the latter alternative. The purpose was to generalise from narrow to broad – from the 53 DF participants to the DF discipline. There are several threats to external validity, which means there are reasons why the inferences about generalisability may be incorrect. These threats relate to the interaction between the inferred causal relationship with units, outcomes, and settings, over treatment variations, or due to context-dependent mediation (Shadish et al., 2002, p. 87). The aspects relevant to the thesis are discussed below.

Interaction of the causal relationship with *units* means that the effect associated with certain units might not hold if other kinds of units were studied. There is little knowledge about the total DF practitioner population, and it is not possible to determine whether the characteristics of DF practitioners participating in the DF experiment correspond with the overall DF practitioner population. Therefore, transparency about the background variables was essential to enable an assessment of transferability to a defined population of DF practitioners.

Interaction of the causal relationship over *treatment variations* entails that an effect found with one treatment (the context) might not hold with other treatment variations, in combinations with other treatments, and when only parts of the treatment are used. This is a central aspect of the DF experiment, since the research has shown that the DF practitioner would often receive contextual information in submission forms or through dialogue with

criminal detectives in an actual criminal investigation setting. They would probably thus have access to more information and information about other evidence types in contrast to the limited information they received in the DF experiment. Hence, it is not possible to predict whether and how strongly contextual information would influence the analysis performed in an actual investigation.

Another aspect is the role of the context as *treatment*. The participants were handed the same evidence file, which contained much information, most of which was irrelevant to the case under investigation. This information may be seen as additional biasing context, which may have created noise or influenced the result. Also, since the DF experiment was conducted over a period, it is possible that DF participants talked about their experience to participants who had not yet completed the experiment, despite the request not to share information, and thus introduced expectancies concerning the scenario, the context, or what traces they would find. Such treatment variations could not be fully controlled, since participants were allowed to perform the experiment at their own lab and on a date of their own choice.

Interaction with the causal relationship with *settings* is about whether the cause-effect relationship between context and biasing effect holds in another setting. Since this is the first experiment specifically targeting bias in DF investigative work, it is uncertain whether the effect may relate to the particular case type and the particular setting introduced in the DF experiment design. DF practitioners are assigned to a plethora of case types, some of which may include content that is more emotionally or morally provocative than others. The case type used in the DF experiment was relatively neutral, and other case types could potentially lead to other or stronger effects due to different settings.

Context-dependent mediation has similarities with the above-mentioned interaction related to settings but centres on mediating processes. The studies of causal mediation identify the essential processes that *must* occur to transfer an effect. Still, the correct mediator may be context-dependent and mediate the effect in one setting and not the other. The contextual bias has been explored in the typical tasks of several other forensic science disciplines, and a biasing effect has been determined in various decision-making tasks such as observations and conclusions (Cooper & Meterko, 2019; Kukucka & Dror, 2022). The DF experiment indicates that the mediator (contextual information) also may introduce contextual bias in DF casework.

Ethical considerations

In addition to the considerations concerning mock or real-life evidence files, as outlined in section 4.1.2 b), an issue concerning volunteer participation and anonymity emerged. A core ethical aspect of research is that participation should be based on informed consent and anonymous participation. The DF practitioners were informed about the purpose of the research, how data would be stored and used, and that they could withdraw at any time (see Appendix 5). Since the contact information of DF practitioners in law enforcement was not publicly available, it was necessary to cooperate with the management of the respective organisation to reach the potential participants. Most managers solved this by sending a list of potential participants and the opportunity to contact them directly. However, some managers seemed to see the DF experiment as an opportunity to gain better insight into the performance of their employees and asked for insight into their employees' results. This request was denied because it would compromise the anonymity of the participants, and there was a risk that unfortunate results (from the manager's perspective) could result in consequences for the participants. Moreover, it would introduce a possible confounding variable that could create noise in the results. The managers who took this approach did not return contact information of any prospect participants, and it is thus likely that the decision to deny access to non-anonymised results led to fewer participants recruited to the DF experiment.

4.4 The professional position and scientific worldview

Philosophers of science such as Thomas S. Kuhn and Jürgen Habermas underline that all research is, to some extent, theory laden. The researcher is not a neutral instrument for presenting facts but brings their knowledge, professional experience, and personal habitus. Hence, it is important to consider how my background and experience may have influenced the research presented in the thesis. I am a sworn police officer with 22 year's experience from the Norwegian police. My experience is largely criminal investigation, spanning from high volume crimes to specialist areas such as homicide investigation and cybercrime. Since joining the Norwegian Police University College in 2012 the role changed from practitioner to 'pracademic', where I have lectured bachelor's, master's and post graduate students in subjects related to criminal investigation and DF investigations.

The academic journey for exploring cognitive and human factors in DF work started with my master's thesis, where DF practitioners and criminal detectives from the Oslo Police District were interviewed about how they planned, conducted, and collaborated in investigations involving digital evidence. This study led to several new insights about the competence

needed to handle digital evidence, issues concerning collaboration, information sharing, and the risk of bias and organisational challenges such as prioritisation and case management. The master's thesis paved the way to the research question of the PhD thesis. Due to my background as a sworn police officer, I am what Brown (1996, p. 181) describes as an "outside insider" researcher. I consider the research objective a product of my knowledge about and experience with the inner workings of law enforcement handling digital evidence. An important issue is whether this position affected my ability to reflect critically on my own profession due to solidarity and a wish to tone down unpleasant findings. From my point of view, the years at the Norwegian Police University College have created a distance to the field, which has helped me obtain more of an outsider's perspective of the DF discipline and broaden my perspective outside the Norwegian context.

As outlined in the introduction (section 1) and as shown in the discussion of theoretical perspectives (section 3), the PhD project is interdisciplinary – understood as "integrated perspectives from different disciplines that add up to more than the sum of their parts" (Silvast & Foulds, 2022, p. 10), by drawing on research from different scholarly traditions. A *mixed methods approach* (Heap & Waters, 2019, p. 10) was applied to explore the research question from different angles. The approach involved a combination of qualitative, quantitative, and experimental methods, which are founded on different worldviews. Creswell and Plano Clark (2017, pp. 63-65) discuss whether one should seek to find the best-suited approach for the chosen mixed methods design or pragmatically choose what fits the particular study. From the start of the project, the choice of methods were based on a consideration of what would potentially provide the best data to examine the research question, and provide a foundation to explore the different perspectives articulated in the research questions. Conflicting positions introduce a challenge when the results are to be integrated and discussed in context. The experiment is rooted in a stance leaning towards the positivist philosophical tradition with an objectivist view of reality, while the questionnaire asking open-ended questions about how the participants handled the physical and cognitive tasks during the experiment belongs to an interpretivist/constructivist tradition founded on a subjectivist view of reality. My ontological position is thus intermediate, and I acknowledge that both objective and subjective views of the reality are useful in the study of DF practices. The differing worldviews would introduce substantial challenges to the integration of results if they were understood from a conservative point of view. Yet, the worldviews seem to be perceived as more flexible in applied research. Today, it is generally acknowledged that

experiments are theory laden and open to multiple interpretations, which may be affected by the researcher's own beliefs, hopes, expectations, and predictions, and the criticisms are often overgeneralised (Shadish et al., 2002, p. 460). The DF experiment is, in essence, a study of socio-technical practice. I consider my own worldview closest to the interpretivist stance, and acknowledge subjectivity and interpretation not only influences the research but also is an important tool for providing insights about DF practices. Subjectivity and interpretation are acknowledged from the researchers' perspective and are also at the core of what is explored concerning the research objects, the DF practitioners.

5. Summary and integration of results

The research has provided novel insights into the hidden or black-boxed practices in the analysis and presentation stages of the DF process.

Article 1 addresses the thesis's primary research objective from a theoretical point of view and contributes to a broader understanding of the challenges related to the DF process of non-technical type, by bringing in theoretical aspects and research on cognitive and human factors from other forensic science disciplines and relating the findings to the DF discipline. The research question of **Article 1** was "*When handling digital evidence through the digital forensics process; when is the digital forensics practitioner (DFP) vulnerable to cognitive bias, and what measures could be relevant and effective to mitigate bias for this specific domain?*" The paper reviewed relevant research on the contextual bias from forensic science domains such as DNA, fingerprint, bloodstain pattern, arson investigation, forensic pathology, forensic anthropology, and crime scene investigations. It aimed to bring these insights into the DF discipline and discuss their relevance and plausible implications for DF casework. A theoretical analysis of the DF process was performed based on a taxonomy of sources for cognitive bias (Dror, 2017). The analysis suggested that the DF practitioner plays an essential role in all stages of the DF process, which involves a magnitude of observations, interpretations, judgements, and decisions. The paper suggests that all the taxonomy levels constitute biasing sources for DF work. In light of the lack of standardisation and quality management in the DF domain, the article emphasises a concern about cognitive and human error leading to miscarriages of justice. The article concludes that there is a research gap concerning bias and DF decision-making and asks whether the suggested bias minimising countermeasures suggested for other forensic science disciplines are relevant and effective in DF investigative work. The paper pointed to the Hierarchy of Expert Performance (HEP) as a possible framework for further research on bias and reliability in DF decision-making.

The value of the research was highlighted in the editorial in the journal *Digital Investigation*, titled Maturation of digital forensics:

The treatment of cognitive and human factors in digital forensics by Nina Sunde (Norwegian Police University College) and Itiel Dror (University College London) takes a major step towards integrating digital forensics with forensic science. This work builds on the strong base of Sunde's study of non-technical sources of errors in

digital investigations, and Dror's extensive experience with cognitive biases and human factors in forensic science. Everyone who employs digital forensics has a responsibility to implement strategies that mitigate those influences which “might interfere with accurate observations and inferences in forensic decision making.” There is a pressing need to update digital forensic effective practice guidelines, error mitigation strategies, and associated training to include bias mitigation practices. (Casey, 2019b, p. A1)

The article prepared the ground for the DF experiment and **Article 2**, which – based on the HEP framework (Dror, 2016) – explored the following research questions: First, “*Are DF examiners biased by contextual information when making observations, interpretations of observations, or in their conclusions during the analysis of digital traces?*” and, second, “*Are DF examiners consistent with one another when making observations, interpretations of observations, or conclusions during the analysis of digital traces?*” While **Article 1** analysed the complete DF process, **Article 2** centred on the analysis and presentation stages (see description of stages in section 2.3). The experiment design aimed to achieve high ecological validity by replicating a typical DF investigation, as opposed to performing the experiment within a highly controlled laboratory environment. Fifty-three DF practitioners participated in the experiment. The background study of submission forms and commissioning procedures showed no restrictions to including task-irrelevant information in the forms. It was thus likely that task-irrelevant information could be forwarded to the DF practitioner in a typical work situation.

In terms of the first research question, the participants were assigned to groups and were given different contextual information. The results showed statistically significant variations between the experiment groups in the number of traces they found. This finding was important, since it first showed that DF is not exceptional among other forensic science disciplines regarding contextual bias. Second, it provided insight into how early in the analysis process a bias may influence the results. The study shows that bias can skew what is observed and deemed relevant to include in the analysis report – and, consequently, that relevant information may be overlooked. From a quality management perspective, this finding is vital. In DF, the report may be subject to quality control, but the evidence file is rarely examined a second time. Verification procedures are directed towards verifying positive findings and not to checking whether there is other relevant information (inculpatory or

exculpatory) that should be included as evidence. As of now, only the top-level “Re-examination” in the Peer Review Hierarchy for DF (Horsman & Sunde, 2020, p. 8; Sunde & Horsman, 2021, p. 9) would be an adequate measure to uncover overlooked relevant evidence caused by a contextual bias.

The findings regarding the second research question were probably the most extraordinary. Here, the reliability between DF practitioners within the groups that received similar contextual information was explored. The results showed low reliability at all explored levels, namely, observations, interpretations, and conclusions. The high variability displays the significance of the human element in DF examinations and is a plausible consequence of the lack of standardisation of investigation methods within the DF domain. In light of the widespread misconceptions of digital information as credible and objective sources of evidential information, this finding is vital. The results underpin the fact that DF work is a highly constructive enterprise in which the DF practitioner significantly influences the outcome of the DF process – the digital evidence.

The novelty in **Article 2** is twofold. First, the existing empirical research on DF decision-making is minimal, and this is the first study that specifically targets the issues of bias and reliability in DF decision-making. Second, **Article 2** also represents methodological novelty, due to the unique research design for studying DF practice. The empirical material in other studies is mainly collected through ethnographic approaches and interviews. Except for the NIST Black-box study (2022), the few experimental studies targeting DF practitioner conduct and decision-making have been performed in laboratory settings. In contrast, the DF experiment aimed to replicate a typical work situation. Also, asking the participants to write an analysis report and using this as an empirical source for the research is a novel approach to studying DF practice.

The statistical measurements in **Article 2** are essential. At the same time, they are limited to what may be counted and do not reveal *how* the contextual bias or low reliability materialised itself in the reports. **Article 3** applied a qualitative lens to the reporting practices during the presentation stage of the DF process. The analysis focused on descriptions of digital evidence. The concept “interpretative flexibility” (Collins, 1981, p. 4; Doherty et al., 2006, p. 569), combined with the components that make up the evidence value (relevance, credibility, inferential/probative force or weight of evidence), was applied as an analytical framework. The article shows that the same traces are interpreted and described differently (a “digital evidence multiple”), and it develops the concept of “evidence elasticity”. It also demonstrates

how the interpretative flexibility provides elasticity when constructing narratives, enabling multiple narratives about what happened and who was involved, based on traces from the same evidence file. The ground truth was known, since a mock and well-documented evidence file was used. Hence, it was possible to explore whether some DF practitioners produced misleading or erroneous descriptions or conclusions. Nevertheless, the analysis showed that many descriptions were vague, slightly inaccurate, or incomplete. This ambiguity enabled the same trace to fit just as nicely with, for example, a narrative imposing guilt as a narrative indicating the innocence of the suspect. These insights are of great value to the DF discipline. They may also be of value to other forensic science disciplines dealing with activity level issues, such as forensic pathology and crime scene investigation, since they shed light on the range of interpretative flexibility for the individual traces or traces in context. Finally, the article demonstrates that interpretative flexibility relates to evidence value, and that traces and narratives may be crafted to highlight the *relevance* of particular evidence, inflate or deflate their *credibility*, and implicitly or explicitly convey opinions concerning the *inferential/probative force or weight* of evidence.

While **Articles 2, 3 and 5** used the DF practitioners' analysis reports as the empirical foundation, **Article 4** was based on their own accounts. The empirical data were collected through a survey, which the DF practitioners completed just after concluding the analysis in the DF experiment. The responses describe their investigative practice during the analysis stage, independent of what they documented in the reports. **Article 4** examined the research question: "*How do DF practitioners handle contextual information, approach examiner objectivity and evidence reliability during the analysis of an evidence file?*" Whist **Article 2** showed that the observations were biased by contextual information, this material provided insight into *how* this may have happened. An important finding in this respect is that many considered hypotheses after reading the scenario information. A hypothesis-driven approach is a well-known bias-minimising measure in criminal investigations and was, therefore, an interesting finding in a DF setting. Yet, some critical aspects concerning how they applied the hypotheses and handled the contextual information were discovered: First, many started working with a single hypothesis in mind, which introduces the risk of a one-sided investigation. Second, considering the possibility of innocence and actively looking for traces that correspond with this hypothesis is vital for safeguarding the presumption of innocence during the analysis. Among those who had a hypothesis, only 55% of these had an innocence

hypothesis, which entails that 45% started the analysis without an innocence hypothesis in mind.

In terms of examiner objectivity, 34% applied no techniques to maintain their objectivity during the examination of the evidence file, which is of concern considering the results from **Article 2** that suggested bias at the observation level. The hypothesis-driven approach was the most frequently mentioned technique among those with applied techniques. However, the DF practitioners generally referred to this technique as a mental process or thinking aid, instead of a structured hypothesis-driven approach where the hypotheses are written down and tested systematically. In light of the biased observations found in **Article 2**, these findings indicate that merely thinking about hypotheses is not a sufficient measure to minimise bias and safeguard fair investigation. Future research should focus on whether a structured hypothesis-driven approach with systematic and documented hypothesis testing would mitigate bias in DF casework.

Considering evidence reliability, 38% did not use any techniques to examine or control the reliability of the uncovered traces. This finding is of concern when considering the risk of technical errors, such as software implementation errors and the risk of misinterpretation by the DF practitioner. The finding is also relevant in the context of **Article 5**, which revealed substantial deficiencies in the documentation practices concerning the application of such techniques.

While other studies of DF practice have focused on the tangible physical tasks or social aspects of DF casework, there is minimal research on DF decision-making during the investigative tasks in the analysis and presentation stages. **Article 4** contributes to opening the black box of DF casework and informs on the DF investigative and cognitive practices that are usually hidden or at least not readily observable by others. It adds to the research on forensic confirmation bias, by shedding light on how the contextual information is used to generate hypotheses when approaching an evidence file. A strength of the paper is that the responses were obtained immediately after performing the analysis and related directly to the task. This approach is likely to have rendered more accurate information about what they *actually did*, rather than what they *ought to do* according to procedures and guidelines.

Article 5 examined DF reporting and documentation practice and stated the following research questions: *“First, what characterises the opinions used in DF reports in terms of opinion type, uncertainty expressions, addressed issues, and included content relevant to the*

credibility assessment of the reported results? Second, to what degree does the DF reporting practice concerning applied opinion types and included content deviate from other FS (forensic science) disciplines?" Three themes from the reports were examined, to address the first research question: First, how conclusions were articulated and which level (source, activity, offence) they addressed; second, the inclusion of content with relevance to the credibility of the information presented in the report; and third, how (un)certainty expressions were articulated.

The analysis showed that *categorical conclusions* or *strength of support* conclusion types typically were applied. While many claimed to have focused on multiple hypotheses during their examinations according to the Part 2 survey (see **Article 4**), there were few traces of these in the reports. Instead, the analysis revealed a strong tendency to present evidence related to a single explanation. This finding corresponds well with the accounts from the DF practitioners regarding using hypotheses as a mental approach to safeguard examiner objectivity during the analysis. **Article 5** revealed another critical aspect related to contextual bias. The biasing contextual information was rarely documented in the analysis reports. It would thus not be available for scrutiny from peers or legal decision makers.

The analysis of content showed substantial deficiencies in the reporting practices, and that essential information often was missing or vaguely described, such as method descriptions and statements about the reliability/validity of methods or tools. These challenges did not seem exclusive to the DF discipline since similar tendencies were also found within the forensic science disciplines used for comparison.

A plethora of (un)certainty expressions was used but without a corresponding explanation of meaning or reference to an established framework. The different articulations resulting from the DF practitioners' subjective assessment of (un)certainty may also increase the interpretative flexibility of the conclusion for those reading the report.

To address the second research question, the findings concerning DF practice were compared with a study involving eight other forensic disciplines. The comparative analysis showed significant differences in conclusion types and report content. Yet, the deficiencies concerning report content seemed to be a common challenge to the DF discipline and the compared forensic science disciplines.

Whilst there is a growing body of academic papers discussing how to evaluate evidence and report the results from a DF investigation (see section 2.3.3 and 3.2.1), there is little to no

research on DF practitioners' practices concerning these issues. This study provides the first empirical account of what characterises DF opinions and conclusions at an investigative stage and informs on how the investigative process is documented. The comparison with other forensic disciplines is also of value, due to the ongoing efforts to harmonise the DF discipline with the forensic science domain, as described in section 2.1. Article 5 was elected for the PW Allen Award for the most meritorious paper published in Science & Justice in 2021 (Appendix 16).

When considering all the five articles in context, the contribution of this research is substantial. The papers offer novel insights that inform the knowledge gap concerning DF investigative and reporting practice. They inform on the significance of the human's role in crafting digital evidence through the DF process. Having insights into how the human influences what is discovered, how it is interpreted, and the notion of evidence credibility is vital to understand how errors and misinformation may occur, and to design adequate investigative procedures and quality measures. From a more general forensic science perspective, the research is valuable with respect to strengthening the scientific foundation within the DF domain. The novel insights into investigative tasks and reporting practices are valuable to the efforts to harmonise with other forensic science disciplines, since they shed light on what is common to other forensic science disciplines vs what is unique to the DF discipline. In particular, the ability of digital traces to inform issues at activity and offence (intent) levels seems to be particular to the DF discipline, and more research is needed to establish how these aspects should be adequately reported.

6. Concluding analysis and discussion – implication of findings

Before the concluding analysis and discussion of the thesis's contribution, the main research question is repeated: *How could a better understanding of the DF practitioner's role in constructing digital evidence within a criminal investigation enable mitigation of errors and safeguard the fair administration of justice?*

The thesis's contribution may be condensed into three key points. First, the thesis contributes to a broader understanding of the mutable components of digital evidence, which will be discussed in section 6.1. Second, the thesis expands the insights concerning the DF practitioner's role in constructing and negotiating the mutability of the evidence. This is debated in section 6.2, in light of different perspectives of investigative work outlined in section 3.2. Third, the thesis has provided novel insights into cognitive and human factors – and how they may cause or contribute to errors of justice. Section 6.3 discusses the implications of these findings and possible measures to mitigate error and misleading results due to bias and noise.

6.1 A broader understanding of the mutable components of the digital evidence

The thesis's contribution has advanced the insights into what makes the digital trace become meaningful digital evidence with a definable value in a legal context. The DF literature has conceptualised the physical or tangible aspects that may influence the evidential value of digital traces as “evidence dynamics” (Casey, 2011b, p. 27). The theoretical discussions concerning error and error mitigation in the DF scholarly literature have mainly been limited to preventing evidence dynamics, and the aspects involving human interpretation and representation have not been subject to substantial attention from researchers. A plausible explanation relates to the discipline's academic roots. DF is primarily considered a specialisation of computing, with its academic home in computer science and computer/software engineering (Jordaan, 2021, p. 6). These domains focus their research on information systems and computers. Combining and integrating theoretical concepts and analytical perspectives from DF, forensic science and cognitive psychology with social science perspectives enabled the examination of another dimension of dynamics or mutability related to digital traces, namely, the interpretative flexibility of digital traces. These insights were advanced further by analysing their mutability in light of legal criteria assessing evidence value. In order to clarify the difference from the physical or digital aspects encompassed by the term “evidence dynamics”, the interpretative flexibility related to digital

traces was termed “evidence elasticity” in **Article 3**. The concept of evidence elasticity moves the focus from digital evidence as merely objects, to emphasise the subjective perception of the knowledge represented by the object. After being interpreted and inscribed in reports by the DF practitioner, the description of the trace becomes stable. Yet, the thesis has demonstrated that different DF practitioners produce very divergent descriptions. Due to evidence elasticity, these may sometimes be accurate and correct representations of the traces – but sometimes also misleading or incorrect.

Interpretative flexibility is not a novel concept in STS research and has revolved around the heterogeneous ways people understand the same object. Star and Griesemer refer to the different perceptions or interpretations as “plasticity” (1989, p. 397) and the objects having this trait as “boundary objects” (1989, p. 387). Still, mutability has primarily been explored in relation to moving knowledge objects between different or epistemic cultures. Of particular relevance is Kruse’s (2016) study of forensic evidence’s movements between the different epistemic cultures in the justice chain and Dahl’s (2009) study of similar aspects concerning DNA evidence. Whilst movement has not been the primary focus of the thesis, it advances insights into the *range* of the elasticity of digital traces in isolation, and when combined when the interpretation is performed by what is assumed to be a single epistemic culture, the DF practitioners.

The experimental research design enabled a novel insight into the *relative range* of the elasticity of the observed traces – which would not be possible through a design with fixed multiple-choice output, such as, e.g., the NIST black-box study (Guttman et al., 2022). The DF practitioners were not asked to find or write anything in particular during the DF experiment, and decided themselves what traces were relevant and how to describe them in their reports. This resulted in rich qualitative material, which enabled the examination of evidence elasticity combined with other qualitative concepts, such as the credentials for evidential value from a legal context.

The range of elasticity concerning digital traces is black-boxed in real-life casework through the inscription process. The result becomes a stable representation of the trace and the related inferences in the written analysis report. This process turns them into what in ANT is conceptualised as “immutable mobiles”, which can move in time and space more or less unchanged (Jonsson & Holmström, 2005; Latour, 1987). At the same time, the report

functions as an interpretative limitation (Santos, 2014, p. 191), since it represents only one of the multiple versions of the traces. Although the research for the thesis did not follow the further movement of the DF reports, it is possible to make a few plausible predictions concerning the implications of the stability established by inscribing the results into the analysis report in light of the findings concerning the relative range of elasticity. The descriptions of evidence in reports are stable, but the knowledge provided by the written representation is still elastic, for several reasons. First, due to the perception and subjectivity involved when the legal decision maker – often with little technological expertise – tries to understand what the evidence is, what it means, and its value in terms of relevance, credibility, and inferential force/weight to the case under investigation. Second, because the case circumstances it is assessed against may change. Third, since the DF practitioner may be called to court to present and explain the evidence, the oral presentation may thus adjust or add information to the knowledge represented by and associated with the trace. Therefore, it is plausible that those reading the DF reports or listening to an oral presentation of it, such as criminal detectives, prosecutors, and judges, would also interpret the same evidence differently.

Summarised, the insights concerning the relative range of the elasticity of digital traces contradict the beliefs that digital evidence should be presumed to be objective, value-neutral, and reliable evidence. The thesis shows that digital evidence should be subject to scrutiny, not only to evaluate evidence dynamics but also for a critical assessment of evidence elasticity, to prevent erroneous or misleading representations of the trace causing errors of justice during the investigation or trial.

6.2 The DF practitioner's role as mediator in a technosocial process from trace to evidence

While the former section discussed the mutable components of the digital traces, the focal point here moves to the role of the DF practitioner in mediating these components. First, the DF practitioner's investigative practice is discussed in light of the perspectives of investigative work outlined in section 3.2. Then, the inscription power of the DF practitioner as an obligatory passage point is debated, followed by a discussion of how the DF practitioner may mediate the evidential aspects of relevance, credibility, and

inferential/probative force or weight. Finally, the implications of the findings from the DF experiment concerning contextual bias and reliability are discussed.

6.2.1 DF investigation – science, investigation, or innovation?

Based on the DF experiment’s empirical foundation, it is not possible to provide generalised knowledge about whether and to what extent the DF practitioner performs the investigative task as a scientific inquiry, that is, if they can apply “science to determine facts” (M. S. Olivier, 2016a, p. 47), or use scientifically derived and proven methods to identify, analyse, interpret, document, and present digital evidence (Årnes, 2018, p. 4). Signs of such an approach would, for example, be descriptions in analysis reports of testing or experimentation to support inferences about a trace, or reports characterised by formalised reasoning, such as the CAI methodology, to produce a preliminary evaluative opinion about the value of the trace to a defined set of hypotheses/propositions and conditioning information. Yet, none of the reports indicated that participants carried out scientific experiments or testing related to the DF experiment to support inferences about the trace or followed the formalised methodology for reasoning about the evidence. Instead, as described in **Article 5**, the reports were very diverse and did not point to a standard methodology for performing the analysis or for documenting the process or results. The tools, methods, and procedures were often vaguely described and sometimes not commented on, and none justified the reliability or validity of the applied methods or tools. **Article 4** showed that a high proportion did not apply any measures to safeguard or control evidence reliability, so the reporting and documentation deficiencies discussed in **Article 5** can thus not be explained by poor documentation of methods and procedures that were actually applied.

Due to the lack of signs of scientific inquiry in the collected reports, it is relevant to explore what would be a more proper characterisation of the observed DF investigative work. DF practitioners are expected to handle a broad range of technologies, and Ward’s (2021) research highlighted the necessity of improvisational skills to obtain information from digital devices such as mobile phones. It could thus be argued that DF investigation may be characterised by ad hoc *innovation* rather than applying a rigorous scientific approach to solving urgent challenges in the DF process. Innovation is described as “new applications of knowledge, ideas, or methods which generate new capabilities and leverage competitive sustainability” (Schniederjans & Schniederjans, 2015, p. 2). Programming and scripting are regarded as essential DF practitioner skills by many academics, students, and professionals in the DF domain (Humphries, 2019, pp. 187-188, 266, 340). Such skills make them – at least in

the eyes of those with less expertise in computers and technology – *a jack of all trades*, with a positive connotation.

In light of the thesis' research question, the ability to apply innovative approaches to obtain digital traces may lead to the closing of information gaps and a richer and more complete empirical foundation for drawing inferences about the questioned matters in the investigation and may thus be regarded as a utility. On the flip side, a DF process characterised by innovation may put the development of methods and tools for obtaining relevant information at the centre of attention, leaving less room for establishing best practices for securing, analysing, interpreting, evaluating, and presenting results, or quality control measures for minimising errors and misinformation. The demand for novel methods to solve the case-by-case challenges may also hinder the DF practitioner from gaining necessary in-depth expertise concerning the technology source in question and from performing the necessary testing and validation of the improvised approaches and methods. If so, there is a risk that the DF practitioner becomes the less flattering extension of the popular quote stated above – the *master of none*.

6.2.2 Mediating relevance, credibility, and evidential value through inscriptions

At their lowest abstraction level, digital traces are binary numbers, which may not be directly observed in a meaningful way. Hardware, software, and various technological instruments are necessary not only to collect but also to observe and assess the relevance of the traces. In order to make sense as meaningful pieces of evidence, the data must be represented at a semantic level, in which the DF practitioner plays an essential role. The thesis has provided new and valuable insights into the DF practitioners' inscription power, due to their exclusive opportunity to frame and shape the digital traces as knowledge objects. Since they possess the necessary combination of expertise and access to technology, they become obligatory passage points in the journey from trace to evidence. The inscription power and their role as obligatory passage points provide the opportunity to mediate the relevance, credibility, and value of the evidence and are discussed below. This is followed by a discussion of the implications of the findings concerning the contextual bias and the low reliability.

The evidence file collected from, for example, a smartphone would usually contain some relevant information combined with a magnitude of information irrelevant to the case under investigation and the task assigned to the DF practitioner. Since only relevant information should be used as evidence, it is not the entire evidence file that is represented in analysis reports and presented in court but the *outcome* of the analysis and the selected traces. The

traces' relevance is mediated in a technosocial process involving the DF practitioner, software, hardware, and the normative and cultural guidance for conducting the process and presenting the result. The DF practitioner may – intentionally or unintentionally – mediate the relevance in several ways. Central to this mediation is the process of inscription in analysis reports. Similar to the “packaging” of CCTV evidence observed by Brookman et al. (2021, p. 14), the DF practitioners would usually describe the individual findings in text, tables, visualisations or by cut and paste snippets from the analysis program's representation of the trace. In this lies the power to include, exclude, and mediate the relevance by ordering, highlighting, and downplaying the findings, using rhetorical and visualisation means.

The combined qualitative and quantitative analysis in **Articles 4 and 5** provides insights into how the DF practitioner mediates the credibility of the evidence. The analysis shows that the elements that introduce transparency and auditability concerning how the DF work was performed, such as the applied tools, methods, and procedures, were included to varied degrees in their reports. Without this information, the report reader has no insight into the uncertainties and limitations of the DF process and its result and is left with the choice of whether or not to trust the findings as representations of the truth.

The credibility also relates to whether the DF practitioner has sufficient expertise to perform the task and interpret or evaluate the result. As shown in **Article 5**, the aspect of expertise was rarely documented or commented on in the report. Still, a DF practitioner would sign the report with a work title, which, for many, would be associated with expertise, for example *Special Investigator*. The background information collected in the Part 1 survey revealed a broad range of education and experience behind such titles, which are not related to a defined and generally accepted level of expertise. The various titles that suggest associations to expertise, combined with missing information on the DF practitioners' background, may create a pretence of expertise, whilst the practitioners, in fact, may not be sufficiently qualified to perform the assigned task.

The components for demonstrating credibility are not only a matter of including or excluding information about expertise and methods/tools or procedures but also a matter of clarity. As shown in **Article 5**, sometimes the reports included these components, but the descriptions were vague and insufficient to fulfil the purpose of transparency and auditability. A paradoxical function of the vague descriptions might be that they cultivate a *notion of credibility* for an unskilled report reader. The notion of credibility may also be mediated by technical expressions and jargon particular to DF work. Santos (2014, p. 200) states that the

epistemic distancing of forensic labs appears in the dimensions of purification, classification, and interpretative limitation, to preserve institutional and scientific credibility. In the context of DF, the technical terms and jargon may be a form of epistemic distancing from the general investigation. The missing, incomplete, or vague information about expertise, methods, tools, and procedures can be signs of what Costa and Santos (2019, p. 472) refer to as a “bubble culture”, characterised by defensive attitudes when reporting and testifying, to maintain the shield of neutrality.

The summary or conclusion was an important part of the DF reports, since this is where the discovered traces’ evidential value to the case under investigation are disseminated. The findings were combined into plausible narratives or sub-narratives about what had happened, how it was performed, who was involved, and their possible intentions. The summary/conclusion was often placed at the beginning of the report, meaning that the reader would be presented with the narrative before the individual findings. The plausible and coherent narratives were usually characterised by one or more reservations, indicating that the narratives explaining the traces were associated with uncertainty. The reports did not show any coherent way of expressing (un)certainty, which increases the interpretative flexibility of the evidential value when the legal decision maker assesses the reported findings.

As described, the conclusion was often framed as a single narrative. This way of presenting the evidence is a way of mediating the value of evidence in a certain direction, for example suggesting criminal guilt or innocence. A coherent and plausible narrative may be convincing, and the traces appear to have probative value for the case. Still, if the same traces can explain a different narrative, they are, in fact, neutral in terms of evidential value. For example, the fact that a spreadsheet was sent from the suspect Jean’s laptop may appear to be valuable information for a narrative suggesting that Jean leaked confidential information to someone unauthorised. Nevertheless, there were more plausible narratives based on the available traces on the evidence file. The traces also corresponded with a scenario involving someone hacking the system and taking over Jean’s account and sending the spreadsheet to an email account they controlled, or a scenario involving another employee accessing Jean’s laptop when unlocked and sending the spreadsheet to their private email account. Summarising the findings with a single narrative is thus a powerful means to mediate the value of the evidence or conceal that it is, in fact, neutral.

6.2.3 The function of the mutable components for evidence and narrative crafting

The thesis suggests that a significant trait of digital evidence, compared to other types of forensic evidence, is its ability to contribute to the various components that make up a narrative. Such components would typically be a scene, motive, actor, and consequences (Pennington & Hastie, 1993). Forensic science disciplines dealing with physical traces are, first and foremost, able to inform source relationships, which involves attributing traces to individuals, for example through DNA analysis and fingerprints, or classifying and comparing materials, such as trace evidence, marks, body fluids, and drugs. Some forensic science disciplines such as bloodstain pattern analysis, arson investigations, or forensic pathology may inform issues or narrative gaps concerning activities or events. Turning to digital evidence, **Articles 3** and **5** shed light on the flexibility of digital traces to inform not only source and activity level issues, but also offence level issues.

Starting with the activity level, digital traces can provide often very accurate information about activities and events, including issues concerning time and location (physical place/digital space). Such traces can often be attributed to source entities such as user names, email addresses, and chat aliases. However, establishing a connection between entity and person is often complex or even out of reach, based solely on digital evidence. Since multiple people may have access to the same device, determining who used the device at the exact time of interest is complicated. For example, in the evidence file used in the DF experiment, there were multiple active user accounts on the computer. Although someone was using a user account named “Jean”, it was uncertain whether that person was the suspect, Jean Jones. Moreover, although a mail was sent from the e-mail account, jean@m57.biz, this does not prove that Jean sent the email. It could, of course, be Jean herself but also someone accessing her computer while she left her desk and forgot to lock it, or someone with unauthorised access to the computer network. Another complicating factor for establishing the source issues is that digital traces always involve computation and often, but not always, human interaction. One must therefore always rule out the trace being machine-generated and justify that it is caused by human activity (e.g., the Trojan defence (Brenner et al., 2004)).

Article 5 demonstrates that digital evidence can inform offence level issues, such as the intent, motivation, planning, or preparations. It is even helpful for classifying the crime (e.g., phishing or unauthorised intrusion into computer systems). The questioned matters of a criminal investigation are rarely examined in the context of the hierarchy of issues but, instead, in the light of the six basic questions of an investigation (T. Cook, 2016, p. 38), often

referred to as 5WH. Digital evidence seems to provide meaningful information for all the basic questions, which are *what, when, where, why, who* and *how*. Turning to the narrative perspective, the narrative's construction is based on the outcome from investigating the issues mentioned above. Traditional forensic evidence has functioned as anchoring points in the narratives, by providing information about who left their DNA trace on the victim, who left their fingerprint on the stolen car, or which weapon may be associated with the bullet cartridge found at the crime scene. Due to their availability and diversity, digital traces are capable of filling the narrative gaps, in terms of both magnitude and detail, or contributing to “mosaicking” (Innes et al., 2021) the narratives with digital traces.

Granhag and Ask (2021, p. 435) outline several essential factors for a narrative to be convincing. The narrative must be plausible and able to explain all or most of the evidence. It must be cohesive and not contradict any evidence. The narrative must be complete, cover all essential parts of the crime, and be arranged in chronological order. Finally, it must be unique, and there should be no alternative narrative with an equal or higher capacity to convince. Costa and Santos (2019) add the presence of forensic evidence to this list of convincing factors. They found that forensic evidence made the narrative appear stronger, even if the evidence was weak (Costa & Santos, 2019, p. 478). The digital traces' ability to fill out the narrative gaps, combined with the elasticity to achieve cohesiveness, completeness, plausibility, chronology, and uniqueness, makes digital evidence a useful means for constructing *convincing narratives*. In addition to the inherent cognitive resistance against questioning what seems to be a plausible and coherent narrative, the digital traces' aura of credibility may add an extra layer of trust and, hence, a cognitive obstacle to scrutinising the individual traces forming in the narrative or the credibility of the narrative as a whole.

To summarise, evidence elasticity adds to the established concept of evidence dynamics, which relates to the physical or digital traits of the trace. In context, these concepts can provide a more complete insight into the aspects that introduce uncertainty regarding digital traces' evidential value and into the role of the DF practitioner as a mediator of evidence relevance, credibility, and probative/inferential force or weight.

6.3 Managing unwanted consequences of elastic digital evidence

The concepts of between-practitioner reliability and biasability make up plausible explanations of the mechanisms that may *stretch* the elastic traces in various directions. While mediation of the evidential components may happen intentionally, the further discussion only

relates to unintentional mediation, which refers to a situation where the DF practitioner aims to be, or believes that they are, an objective and value-neutral analytical instrument in the DF process. This is followed by a discussion about possible approaches to manage low reliability and bias, particularly how to prevent unwanted consequences such as misleading or erroneous results of DF investigations.

6.3.1 Biasability

Due to the substantial body of research from other forensic disciplines indicating that forensic tasks may be biased by contextual information, it was important to explore whether there were similar tendencies in DF work. The results from the DF experiment indicated that DF practitioners are prone to the same tendency as that observed through research in other forensic science disciplines. The DF experiment showed that contextual information, for example provided through submission forms, dialogue about the assignment, or access to the case file, might influence the number of traces the DF practitioners observe and deem relevant to report. Anderson et al. (2005, pp. 58-59) describe the investigative process as “connecting the dots”. Transferring this analogy to the variation in observed traces in the DF experiment, when some relevant “dots” are missing, or non-existing “dots” are included, it could have unfortunate implications for the explanations derived from them.

The DF experiment also examined contextual bias when interpreting and drawing conclusions on the discovered traces but did not show statistically significant results. Nevertheless, it does not rule out contextual bias in DF casework when such tasks are performed, mainly because they are interrelated with the observation level. A contextual bias may cascade, since the uncovered traces are subject to interpretation and later form the basis for a conclusion. Future research on DF practitioner bias should thus include observations, interpretations, and conclusions.

Although it is premature to generalise contextual bias in DF based on a single experiment, it is important to discuss and consider the possible implications of such a bias for DF work and possible measures to minimise effects that may lead to skewed and unfair results of an investigation. Regarding relevance, such measures should aim to prevent one-sided investigations, characterised by case building to substantiate guilt while overlooking and explaining away evidence consistent with innocence. In terms of credibility, measures should aim to prevent asymmetrical scepticism, by ensuring that incriminating evidence is scrutinised just as thoroughly as evidence indicating innocence.

6.3.2 Bias minimising measures

Several bias minimisation measures have been suggested for the forensic science domain, such as context management, compartmentalisation of tasks and techniques, Linear Sequential Unmasking (LSU) (Dror et al., 2015, see **Article 2** for description) and Linear Sequential Unmasking – Expanded (LSU-E) (Dror & Kukucka, 2021). Still, observing similar contextual bias tendencies in DF to those in other forensic science disciplines does not necessarily entail that similar bias minimisation measures are effective, due to differences in the investigative tasks' nature.

Drawing on the research and recommendations from other forensic science disciplines, a plausible measure would be to exclude all task-irrelevant information from the DF practitioner, and keep the task-relevant information away from the DF practitioner until the initial examination of the evidence has been performed, as recommended in the LSU-E procedure. The guiding principle in this procedure is to always begin with the actual evidence, before considering any other contextual information, ensuring that the examiner is allowed to form an initial impression before receiving any biasing contextual information (Dror & Kukucka, 2021, p. 3). However, the procedure may often not be feasible, due to several discipline-specific factors for DF.

First, and as outlined in **Articles 2** and **4**, determining what should be considered a task-irrelevant context is not straightforward at a general level and must be assessed from case to case. Second, a digital device would typically contain a magnitude of data, sometimes even terabytes. Conducting an initial examination of the data without any case knowledge to direct the examiner, as recommended in LSU-E, would often be unachievable, due to the magnitude of data combined with the need for case knowledge. The challenge of exploring all the data on a device may be compared, in a very simplified way, to walking into a library and trying to get an overview of the information there. In the imagined library, some of the shelves are in order, containing organised information. Other shelves are empty, and the books that used to occupy the shelves are found in large piles, among loose pages and pieces of paper, with various content such as text, images, graphs, and signs. Some of the books have their library reference intact and can be placed back on a shelf, while other items lack the reference to the shelf or to which book or text they belong. Walking around to gain a first impression is not meaningful under such conditions. There is too much information to review before receiving any context. Hence, a DF practitioner would be forced to make a choice, and there is no guarantee that the chosen shelves or content from the book piles would contain relevant

information. The reviewed information would become an anchoring point, influencing the subsequent perceptions, interpretations, and conclusions concerning the reviewed data. Third, exploring every bit of the content might not even be a lawful act. Not limiting the violation of privacy to what is necessary and proportionate in light of the investigation's objective could be considered a violation of privacy according to the European Convention on Human Rights (ECHR) Article 8. Based on the outlined issues, the LSU-E is thus considered a feasible measure only when the DF examination involves a small and limited dataset, for example when reviewing limited amount of CCTV footage, which for a typical DF practitioner would be the exception rather than the rule.

For forensic disciplines tasked with causal and process judgements (as opposed to feature comparison judgements), Spellman et al. (2021, p. 13) emphasise that exposure to case-relevant but task-irrelevant context is quite common, due to their investigative role. Although not mentioned by Spellman et al., DF fits well into this category. Hypotheses play an important role in causal and process judgement, but the process of generating and evaluating hypotheses is also prone to bias (Spellman et al., 2021, p. 14). They highlight a context-blind peer review as a possible means for detecting unwanted effects on the outcome due to contextual biases, a measure that is also recommended in the Phase-oriented Advice and Review Structure (PARS) methodology (Sunde & Horsman, 2021, p. 13).

Article 4 suggests that using hypotheses is not unusual in a DF examination. Nevertheless, the DF practitioners' responses showed that they would *think* about hypotheses and not articulate them and test them in a structured manner – in writing. The absence of reference to hypotheses in the reports (**Article 5**) underpinned the fact that they were not used in a systematic manner to analyse the evidence file or disseminate the results. As shown in **Article 5**, the conclusions in the analysis reports were focused on a single explanation, not multiple hypotheses. Evidence presented in light of a single hypothesis may appear to have value for the case. However, if the evidence is, in fact, of equal relevance to both a guilt and an innocence hypothesis – it is of neutral value. Hence, framed in light of a single hypothesis or explanation, the conclusions may have the potential to mislead the legal decision maker.

Based on this, the following recommendations are suggested for DF investigations to minimise bias:

- As a general principle, the DF practitioner's exposure to clearly task-irrelevant information should be avoided, or at least kept at a minimum. The DF practitioner should *explicate in the report the context that was provided/available*. This measure would ensure transparency about what information implicitly or explicitly may have influenced the decision-making during the DF investigation and the interpretation and evaluation of the result, and it would enable scrutiny through quality measures such as peer review.
- If not otherwise provided, the DF practitioner should define a set of hypotheses that should guide the examination based on the assignment. To avoid a guilt bias and safeguard the presumption of innocence, the set of hypotheses guiding the DF investigation should *as a minimum include an innocence hypothesis*. Since both top-down and bottom-up reasoning processes are applied in DF casework, new hypotheses might emerge during the examination, and the set of hypotheses should thus be updated. To use the DF experiment as an example: the task description was "What has happened, and what was Jean's involvement in the reported incident?" In a balanced hypothesis-based assignment, the task could be articulated as follows: Your task is to examine whether H1) Jean was involved in the incident or H2) Jean was unrelated to the incident. Instead of using the hypotheses as a cognitive support, they should be written down and systematically tested against the examined data.
- To avoid one-sided conclusions and misleading presentation of evidential value in DF reports, the *result should be reported in relation to the hypotheses that guided the investigation*. This would ensure transparency concerning the reasoning about the findings and would provide less interpretative flexibility to those that should make use of the results, such as the legal decision makers.

A precondition for the suggested measures is knowledge and awareness of human and cognitive factors. Including in DF education and training topics such as biasing sources, possible cognitive effects and consequences for DF casework and bias mitigation measures would prepare the ground for the successful implementation of the recommended measures in DF casework.

6.3.3 Reliability

Kahneman et al., (2021, p. 33) state: “Wherever there is judgement, there is noise – and more of it than we think”, which is also an accurate reflection of the thesis’s findings. The thesis demonstrates that, in the DF process, the DF practitioner plays an active, significant, and necessary role as an analytical instrument, although not a reliable one – in terms of consistency. The variation is substantial at all the levels explored in the DF experiment: observation of traces, interpretation of the observed traces, and conclusions. After determining the low level of between-practitioner reliability in statistical tests, the variability was explored further through a qualitative lens. The aim was to understand how the low reliability materialised itself in reports and whether and how it could challenge the fair administration of justice.

At the observation level, two types of invalid observations of traces were of particular concern: false positives and false negatives. The false positives entailed that some declared that they had found traces that were *not present* on the evidence file. The false negatives involved some categorically concluding with the absence of traces that were *present* on the evidence file.

At the interpretation level, the interpretations were inconsistent, even for what would require a basic level of expertise. An example of a basic level interpretation in the DF experiment involved comparing the leaked spreadsheet vs a spreadsheet found on the suspect’s computer. As shown in **Article 3**, the interpretations ranged from accurate and correct, partly correct and lacking important information, to incorrect and misleading. This finding is, however, not unique. As described in section 2.3.2, the NIST black-box study found that the proportion that answered basic questions incorrectly about the hard disk image varied from 0-34.3%, and the figures were even higher (0-51.9%) for the mobile phone image (Guttman et al. 2022, pp. 8-10, 22-23). In context, these studies indicate that the variation observed during DF investigation may lead not only to misleading statements in reports but also to erroneous interpretations or conclusions concerning traces, due to unintended practitioner error.

At the conclusion level, the statistical analysis showed that the DF practitioners were inconsistent when assigning value to the individual traces they had found. The qualitative analysis presented in **Article 3** showed that very different scenarios or explanations emerged when the individual traces were joined into narratives or sub-narratives.

For example, one DF practitioner constructed a plausible narrative indicating that the suspect had committed the crime:

If the user account “Jean” on the analysed system is Jean Jones, the findings mentioned here indicate that Jean may have sent the file “m57biz.xls” per email to tuckergorge@gmail.com or alison@m57.biz on the 2008-07-20 03:28:00 CEST. (D42)

Another DF practitioner constructed a plausible narrative involving the suspect being tricked and, thus, innocent:

The examination has uncovered that Jean Jones appears to have been the victim of a spear-phishing attack, where she has sent the leaked document to an attacker, who has pretended to be her colleague, Alison Smith. (D15)

A third DF practitioner concluded that there had been no leakage from the laptop:

Based on the analysis referred to under section 5, no information was found indicating that the spreadsheet, m57biz.xls, or other documents with the same content were shared from the unit. (D28)

Some between-practitioner variation was expected due to the knowledge of the many variables concerning method, tools, and practitioner expertise in DF. Still, the high degree and range of variation was beyond what was expected, probably because it would be black-boxed in actual casework. The fact that the variation involved not only questionable results but also erroneous observations, interpretations, and conclusions is a cause for concern. These insights are vital, since they add to the other well-known technical errors or evidence dynamics that may harm the credibility of digital evidence, as described in section 3.4.1. Understanding the errors and their sources is essential for designing effective preventive measures, as well as measures for detecting errors if they occur despite any preventive efforts.

The magnitude of the problem of low reliability hinges on the ability to detect misleading and erroneous results further up the justice chain. Challenging the applied procedures requires expertise about how they *ought to be* performed to be forensically sound and transparency about how they *were performed* in the questioned case. Probably even more importantly, it would require a belief that such examination could uncover erroneous or misleading results, to find it relevant and necessary to scrutiny procedures and results. The review of the Danish telecom case showed that the criminal detectives did not control the data, despite the

procedure being as simple as counting and comparing rows on two spreadsheets, and the reason seemed to be that they trusted The National Police (Rigspolitiet) to provide them with material of a quality that did not need to be scrutinised (Lentz & Sunde, 2020). Basing thresholds for performing quality control on trust was thus an ineffective strategy.

The thesis contributes to the body of research showing that errors and misleading interpretations can also be produced by experts. This calls for a culture change in DF where assessing the quality of procedures and the accuracy of results is just as important as the examination itself, focusing on what to learn and improve, instead of who to blame.

6.3.4 Noise and variation – friend or foe?

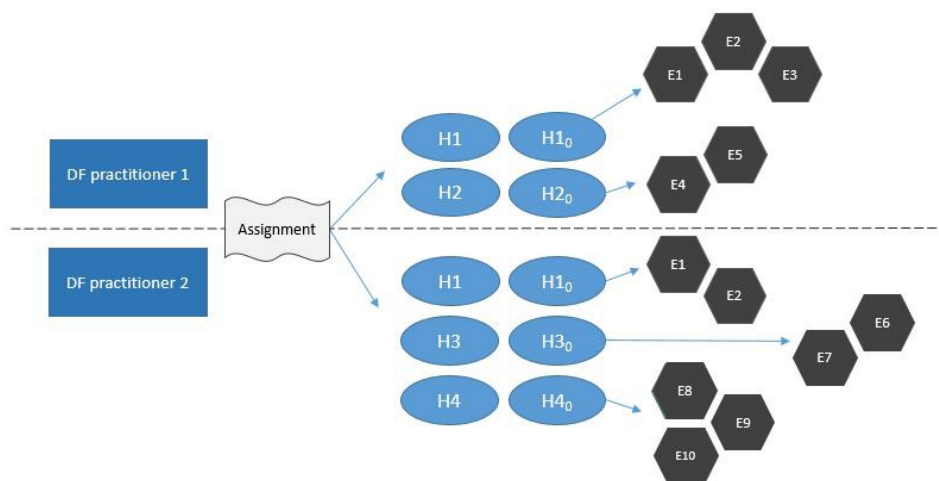
In light of the essential role of the DF practitioner in the DF process and the diverse nature of the scope of a DF assignment, it is very unlikely that we will be able to exclude all noise in the DF investigative decision-making and achieve total consistency in the results. Despite the empirical findings concerning the problematic sides of variation and noise, the opposite perspective should be considered: whether variation may be advantageous to an investigation.

Drawing on quality measures from qualitative research, various forms of triangulation may be helpful to arrive at a fair, well-investigated result, in terms of both relevance and credibility. The logic of triangulation is founded on the premise that “no single method ever adequately solves the problem of rival explanations” (Patton, 1999, p. 1192). Whilst triangulation is often associated with validation by checking the consistency of findings generated by different data collection methods (Heap & Waters, 2019, p. 111), it may also be a means to explore the problem or theme from multiple perspectives (Creswell, 2013, p. 251; Patton, 1999, p. 1193) or by applying multiple methods (Heap & Waters, 2019, p. 112). Applying this view to DF investigation, the aim of triangulation would be to obtain a richer empirical foundation for the results and a deeper insight into the relationship between the investigative hypotheses, the inquiry approach, and findings.

Denzin (1970) distinguished between four forms of triangulation: “Data triangulation”, “investigator triangulation”, “theoretical triangulation”, and “methodological triangulation”. Applying *investigator triangulation*, which involves multiple researchers exploring the same content, is recommended, not only to see if they interpret the same themes consistently but also to utilise their ability to discover various aspects that other researchers may overlook (Denzin, 1970). Triangulation may thus also be useful in a DF setting; however, to avoid bias, the DF practitioners should work independently with the case. Combining investigator

triangulation with *theory triangulation* ensures that the individual DF practitioner develops their investigative hypotheses based on the assignment (see Figure 4). The hypotheses (H1/H1₀-H4/H4₀) form the basis for predicting where relevant traces may be located given a true hypothesis, as well as the choice of method and tool to search for these traces. A combination of investigator triangulation and theory triangulation, as illustrated in Figure 4, may lead to more relevant findings (E1-10) and a more complete empirical foundation for any claims, as well as a more nuanced interpretation of the value of the findings, while minimising any biasing influence from the second DF practitioner.

Figure 4: An illustration of a combined investigator and theory triangulation in a DF investigation.



Applying this view in a criminal investigation context, the biasing sources in the lower levels of Dror's (2017) pyramid, such as personality, experience, education, and training, may be turned into a utility. Different DF practitioners may be able to uncover various traces, due to their previous experience and expertise, and assess the relevance of discovered traces differently, as a result of their knowledge and individual understandings of the evidential themes and information gaps of the case under investigation.

The variation becomes *problematic* only when they disagree about what the exact same traces are, what they mean, or their evidential value in light of identical hypotheses and conditioning information. According to Patton (1999), the key to solving inconsistencies would be to determine the reasons for the differences. To allow scrutiny, the DF practitioners should

therefore be transparent about what hypotheses they formed, the methods applied to test the hypotheses and the reasoning associated with the results.

Adopting the view of noise as a potential utility requires reflexivity of the DF practitioner, acknowledging that there may be multiple interpretations of a single trace, and that the interpretation hinges not only on knowledge but also on applied methods, theories, and perspectives. It requires a view of objectivity as exploring a problem from multiple angles, as opposed to something that is “interpretation-free” (Rønn, 2022, p. 2). Noise will sometimes be the result of different types of expertise, a different perspective, or a valuable extra pair of *tinted glasses*, enabling the detection of traces that complement or nuance, and sometimes also question, the initial findings of the DF analysis. Working this way, the DF practitioner can be critical about their own investigative practice and open to discuss alternative interpretations, in line with the characteristics of the reflexive investigator described by Hald and Rønn (2013, pp. 30-31).

However, the feasibility of such a measure can not be assessed without considering the time and resources aspect. As outlined in section 2.2, the DF discipline is faced with several challenges and a constant struggle to reduce backlogs. The combined investigator and theory triangulation measure will be resource-demanding and should probably be reserved for grave or complex crimes or cold case investigations.

6.3.5 Experiment as a “noise” audit

Despite high ecological validity, an experiment is a setting that deviates from how a DF investigation would be carried out in real life. There would probably never be a case where 53 independent DF practitioners would examine the same evidence file and write individual reports about their findings. Still, an experiment provides unique and valuable insights into aspects of DF work that would be invisible under typical working conditions. In a normal situation, a single DF practitioner would analyse the evidence file alone. Alternatively, they would share the task with a criminal detective, who would perform the content analysis while the DF practitioner performed technical analysis of the identified content. These approaches would not reveal any variation and consequently not initiate a controversy concerning the results. The range of variability would thus not be observable in actual casework through research methods such as ethnographic studies, document reviews, surveys, or interviews.

Kahneman et al. (2021, pp. 370, 379-385) emphasise the importance of gaining insight into situations prone to noise and the range of the noisy decision-making, and that it is possible

through *noise audits*. Such audits make the variation visible, concrete, and tangible, which is a first step towards acknowledging and managing variation. To gain more insight into the degree/amount and range of the variation in the organisations performing DF investigation, the methodology used in the DF experiment could be applied as a quality management measure within DF organisations and units. By engaging several DF practitioners in independently solving the same assignment, the organisation may gain insight into relative variation associated with their typical tasks. Using a mock dataset where the ground truth is known, the organisation may also measure the ability to arrive at valid interpretations and conclusions. A noise audit may thus help gain insight into where the risk of inconsistent decision-making is highest and where preventive measures should be implemented.

7. Summing up the contribution of the thesis

Criminal investigation is one of key strategic functions of police service, and professionalised police practice and investigative quality is vital for police legitimacy (Fahsing, 2016, p. 4; Hestehave, 2021, p. 73; Maguire, 2008, p. 433; National Police Chiefs' Council, 2020). The thesis's objective was to examine how a better understanding of the DF practitioner's role in constructing digital evidence within a criminal investigation could enable the mitigation of errors and safeguard the fair administration of justice.

The thesis points in the direction of changing the way we think about the digital evidence, the DF process, and the DF practitioners conducting the DF investigative work. Instead of perceiving digital evidence as objective, value-neutral, and reliable objects, the thesis suggests that digital evidence should be acknowledged as an elastic form of evidence, whose relevance, credibility, and inferential/probative force and strength is crafted, first and foremost, by DF practitioners. Research involving prosecutors (e.g. Erlandsen, 2019) and cases such as the British Post Office scandal (Flinders, 2021; Virgo, 2021) and the Danish telecom case (Lentz & Sunde, 2020; Sorensen, 2019) show that the wider justice system may have insufficient ability to effectively scrutinise digital evidence and detect any misleading or erroneous claims involving digital traces.

The research offers novel insights into DF practitioner conduct during the analysis and presentation stages in the DF process. Instead of viewing the DF practitioner as a passive operator of various software programs and techniques for identifying undisputable facts, the thesis has shed light on the DF practitioner's significant role as an analytical instrument in the DF process. Expertise, combined with access to the necessary tools, methods, and procedures, makes the DF practitioner an obligatory passage point, with the opportunity to craft the individual digital traces or the narratives they contribute to the formation of. The research has advanced insights into the DF practitioner's inscription power as a means of mediating the evidential value. At the same time, the thesis provides novel insights into the cognitive and human factors that influence DF investigative construction work, and particularly the role of contextual information as a powerful biasing source. Acknowledging that DF practitioners are prone to biases, due, for example, to contextual influences paves the way for research on bias mitigating measures customised for DF investigative work, as well as the effective implementation of such measures. The thesis has also advanced insights into whether DF

practitioners are reliable analytical instruments in the DF process and has indicated low between-practitioner reliability – or noise – in observations, interpretations, and conclusions.

The results indicate that the DF discipline needs to take errors caused by cognitive and human factors into account by advancing the knowledge about the biasing mechanisms, sources, and effects in the DF curricula. Transparency is key to allowing the necessary scrutiny into the possible biasing sources that may have influenced the work, for insight into the applied tools, methods, procedures, and error mitigation strategies in the DF casework. Managing cognitive and human factors effectively involves a view of error as something that should be actively managed in DF casework. Since bias for the most part happens unconsciously, the focus should not be on whom to blame for any mistakes but on how to prevent, detect, and correct systematic errors. The thesis has shown that scrutiny concerning the DF process and the result – the digital evidence – is necessary to prevent errors and misinformation from entering the investigations and safeguard the fair administration of justice.

7.1 Future research

The final section will point towards the most urgent knowledge gaps uncovered through the research. The experiment was the first to explore biasability and reliability in DF decision-making, and a single study is not sufficient to generalise the association between contextual influences and bias in DF work. Context is only one of several biasing sources, and more research is needed on biasing sources and situations that introduce a high risk of biased decisions in DF work.

The thesis explored the analysis and presentation stages of the DF process, with a particular focus on the DF practitioner. The finding concerning low reliability should be followed up with further research, to gain insight into the variation that may exist also in other stages of the DF process. The journey from trace to evidence and particularly how digital evidence is perceived further up the justice chain should be researched further to inform the development of effective reporting strategies.

Digital evidence plays a significant role in solving criminal cases, and the reliance on digital evidence will probably persist, due to the reliance on technology in today's society. Research on investigative practice to underpin effective and fair practices for handling such evidence is thus essential to maintain the trust in law enforcement's ability to perform fair DF

investigations of digital traces and prevent miscarriages of justice caused by misleading or erroneous digital evidence.

References

- Anderson, T., Schum, D., & Twining, W. (2005). *Analysis of evidence*. Cambridge University Press.
- Andersson, O. (2020). *Is the chain unbroken: A pilot study of the local police use of IT forensic processes*. Master's thesis, Halmstad University. DiVA.
<http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Ahh%3Adiva-42551>
- Andreassen, L. E., & Andresen, G. (2020). *Live data forensics: A quantitative study of the Norwegian Police University College students LDF examinations during their year of practice*. Master's thesis, University College Dublin. PIA.
<https://hdl.handle.net/11250/2734964>
- Angius, N., Primiero, G., & Turner, R. (2021). The philosophy of computer science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (spring 2021 edition)*: Metaphysics Research Lab, Stanford University.
- Arntzen, H. H. (2018). *Åsted. På innsiden av Kripas. En kriminalteknikers jakt på sannheten*. Kagge.
- Arscott, E., Morgan, R., Meakin, G., & French, J. (2017). Understanding forensic expert evaluative evidence: A study of the perception of verbal expressions of the strength of evidence. *Science & Justice*, 57(3), 221–227.
<https://doi.org/10.1016/j.scijus.2017.02.002>
- Atkinson, J. S. (2014). Proof is not binary: The pace and complexity of computer systems and the challenges digital evidence poses to the legal system. *Birkbeck Law Review*, 2(2), 245–261.
- Baker, D. W., Brothers, S. I., Geradts, Z. J., Lacey, D. S., Nance, K. L., Ryan, D. J., Sammons, J. E., & Stephenson, P. (2013). Digital evolution: History, challenges and future directions for the digital and multimedia sciences section. In D. H. Ubelaker (Ed.), *Forensic science. Current issues, future directions* (pp. 252–291). Wiley-Blackwell.
- Bali, A. S., Edmond, G., Ballantyne, K. N., Kemp, R. I., & Martire, K. A. (2020). Communicating forensic science opinion: An examination of expert reporting practices. *Science & Justice*, 60(3), 216–224.
<https://doi.org/10.1016/j.scijus.2019.12.005>
- Bali, A. S., Edmond, G., Ballantyne, K. N., Kemp, R. I., & Martire, K. A. (2021). Corrigendum to “Communicating forensic science opinion: An examination of expert

- reporting practices” [*Science & Justice* 60(3) (2020) 216–224]. *Science & Justice*, 61(4), 449–450. <https://doi.org/10.1016/j.scijus.2021.04.001>
- Bayes, T. (1764). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 370–418.
- Bednar, P. M., Katos, V., & Hennell, C. (2008, 9 October 2008). *Cyber-crime investigations: Complex collaborative decision making*. Paper presented at the 2008 Third International Annual Workshop on Digital Forensics and Incident Analysis, Malaga, Spain.
- Beebe, N. (2009). Digital forensic research: The good, the bad and the unaddressed. In G. Peterson & S. Sheno (Eds.), *Advances in digital forensics V, DigitalForensics 2009, IFIP Advances in information and communication technology, vol. 306* (pp. 17-36). Springer. https://doi.org/10.1007/978-3-642-04155-6_2
- Beebe, N. L., & Clark, J. G. (2005). A hierarchical, objectives-based framework for the digital investigations process. *Digital Investigation*, 2(2), 147–167. <https://doi.org/10.1016/j.diin.2005.04.002>
- Bhoedjang, R. A. F., van Ballegooij, A. R., van Beek, H. M. A., van Schie, J. C., Dillema, F. W., van Baar, R. B., Ouwendijk, F. A., & Streppel, M. (2012). Engineering an online computer forensic service. *Digital Investigation*, 9(2), 96–108. <https://doi.org/10.1016/j.diin.2012.10.001>
- Blanco, F. (2017). Cognitive bias. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of animal cognition and behavior* (pp. 1–7). Springer.
- Borhaug, T. S. (2019). *The paradox of automation in digital forensics*. Master's thesis, Norwegian University of Science and Technology. NTNU Open. <http://hdl.handle.net/11250/2617753>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brenner, S. W., Carrier, B., & Henninger, J. (2004). The Trojan horse defense in cybercrime cases. *Santa Clara Computer and High Technology Law Journal*, 21(1), 1–53.
- Brenner, S. W., & Schwerha, J. J. (2002). Transnational evidence gathering and local prosecution of international cybercrime. *John Marshall Journal of Computer & Information Law*, 20(3), 347–395. <https://repository.law.uic.edu/jitpl/vol20/iss3/1>
- Bridger, R. S. (2021). A guide to human factors in accident investigation. In S. O. Johnsen & T. Porathe (Eds.), *Sensemaking in safety critical and complex situations* (pp. 13-32). CRC Press.

- Brodeur, J.-P. (2010). *The policing web*. Oxford University Press.
- Brookman, F., & Jones, H. (2021). Capturing killers: The construction of CCTV evidence during homicide investigations. *Policing and Society*, 32(2), 125–144.
<https://doi.org/10.1080/10439463.2021.1879075>
- Brookman, F., Jones, H., Williams, R., & Fraser, J. (2020a). Crafting credible homicide narratives: Forensic technoscience in contemporary criminal investigations. *Deviant Behavior*, 43(3), 340–366. <https://doi.org/10.1080/01639625.2020.1837692>
- Brookman, F., Jones, H., Williams, R., & Fraser, J. (2020b). Dead reckoning: Unraveling how “homicide” cases travel from crime scene to court using qualitative research methods. *Homicide Studies*, 24(3), 283–306. <https://doi.org/10.1177/1088767920907374>
- Brown, J. (1996). Police research: Some critical issues. In F. Leishman, B. Loveday & S. P. Savage (Eds.), *Core issues in policing* (pp. 178–190). Longman.
- Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, 2(1), 1–38. <https://doi.org/10.5334/joc.72>
- Bushway, S., & Forst, B. (2013). Studying discretion in the processes that generate criminal justice sanctions. *Justice Quarterly*, 30(2), 199–222.
<https://doi.org/10.1080/07418825.2012.682604>
- Böhm, F., Englbrecht, L., Friedl, S., & Pernul, G. (2021, 27 October 2021). *Visual decision-support for live digital forensics*. Paper presented at the 2021 IEEE Symposium on Visualization for Cyber Security (VizSec), New Orleans, LA, USA.
- Callon, M. (1986). Some elements of a sociology of translation: Domestication of scallops and the fishermen of St Brieuc Bay. In J. Law (Ed.), *Power, action and belief: A new sociology of knowledge?* (pp. 196–223). Routledge and Kegan Paul.
- Callon, M. (2001). Actor Network Theory. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 62–66). Pergamon.
- Carlton, G. H. (2007). A grounded theory approach to identifying and measuring forensic data acquisition tasks. *Journal of Digital Forensics, Security and Law*, 2, 35–56.
<https://doi.org/10.15394/jdfsl.2007.1015>
- Carr, S., Piasecki, E., & Gallop, A. (2020). Demonstrating reliability through transparency: A scientific validity framework to assist scientists and lawyers in criminal proceedings. *Forensic Science International*, 308, 110110.
<https://doi.org/10.1016/j.forsciint.2019.110110>

- Carrier, B. (2006). *A hypothesis based approach to digital forensic investigations*. PhD thesis, Purdue University. https://www.cerias.purdue.edu/assets/pdf/bibtex_archive/2006-06.pdf
- Carrier, B., & Spafford, E. H. (2003). Getting physical with the digital investigation process. *International Journal of Digital Evidence*, 2(2), 1–20.
- Carrier, B., & Spafford, E. H. (2004, 11-13 August 2004). *An event-based digital forensic investigation framework*. Paper presented at the Digital Forensic Research Conference DFRWS 2004 USA, Baltimore, MD, USA.
- Casey, E. (2002). Error, uncertainty, and loss in digital evidence. *International Journal of Digital Evidence*, 1(2), 1–45.
- Casey, E. (Ed.). (2004). *Digital evidence and computer crime* (2nd ed.). Academic Press.
- Casey, E. (2011a). Applying forensic science to computers. In E. Casey (Ed.), *Digital evidence and computer crime: Forensic science, computers and the Internet* (3rd ed.) (pp. 465-512). Elsevier.
- Casey, E. (2011b). Foundations of digital forensics. In E. Casey (Ed.), *Digital evidence and computer crime: Forensic science, computers and the Internet* (3rd ed.) (pp. 3-34). Elsevier.
- Casey, E. (2011c). Handling a digital crime scene. In E. Casey (Ed.), *Digital evidence and computer crime: Forensic science, computers and the Internet* (3rd ed.) (pp. 227-254). Elsevier.
- Casey, E. (2013). Editorial: Triage in digital forensics. *Digital Investigation*, 10(2), 85–86. <http://dx.doi.org/10.1016/j.diin.2013.08.001>
- Casey, E. (2016). Editorial: Differentiating the phases of digital investigations. *Digital Investigation*, 19, A1–A3. <https://doi.org/10.1016/j.diin.2016.11.001>
- Casey, E. (2019a). The chequered past and risky future of digital forensics. *Australian Journal of Forensic Sciences*, 51(6), 649–664. <https://doi.org/10.1080/00450618.2018.1554090>
- Casey, E. (2019b). Editorial: Maturation of digital forensics. *Digital Investigation*, 29, A1–A2. <https://doi.org/10.1016/j.diin.2019.05.002>
- Casey, E. (2020a). Editorial: The epic story of scientific interpretation in digital investigations. *Forensic Science International: Digital Investigation*, 34, 301063. <https://doi.org/10.1016/j.fsidi.2020.301063>

- Casey, E. (2020b). Standardization of forming and expressing preliminary evaluative opinions on digital evidence. *Forensic Science International: Digital Investigation*, 32, 200888. <https://doi.org/10.1016/j.fsidi.2019.200888>
- Casey, E., Barnum, S., Griffith, R., Snyder, J., van Beek, H., & Nelson, A. (2017). Advancing coordinated cyber-investigations and tool interoperability using a community developed specification language. *Digital investigation*, 22, 14-45. <https://doi.org/10.1016/j.diin.2017.08.002>
- Casey, E., & Daywalt, C. (2011). Computer intrusions. In E. Casey (Ed.), *Digital evidence and computer crime: Forensic science, computers and the Internet* (3rd ed.) (pp. 369-420). Elsevier.
- Casey, E., & Palmer, G. (2004). The investigative process. In E. Casey (Ed.), *Digital evidence and computer crime* (2nd ed.) (pp. 91-114). Academic Press.
- Casey, E., & Turvey, B. E. (2011). Computer intrusions. In E. Casey (Ed.), *Digital evidence and computer crime: Forensic science, computers and the Internet* (3rd ed.) (pp. 255-284). Elsevier.
- Cervantes Mori, M. D., Kävrestad, J., & Nohlberg, M. (2021). *Success factors and challenges in digital forensics for law enforcement in Sweden*. Paper presented at the 7th International Workshop on Socio-Technical Perspective in IS development, Trento, Italy. <https://www.diva-portal.org/smash/get/diva2:1622611/FULLTEXT01.pdf>
- Charters, I. (2009). *The evolution of digital forensics: Civilizing the cyber frontier*. Unpublished manuscript. <http://www.guerilla-ciso.com/wp-content/uploads/2009/01/the-evolution-of-digital-forensics-ian-charters.pdf>
- Chin, J. M., McFadden, R., & Edmond, G. (2020). Forensic science needs registered reports. *Forensic Science International: Synergy*, 2, 41–45. <https://doi.org/10.1016/j.fsisyn.2019.10.005>
- Chin, J. M., Ribeiro, G., & Rairden, A. (2019). Open forensic science. *Journal of Law and the Biosciences*, 6(1), 255–288. <https://doi.org/10.1093/jlb/lasz009>
- Christensen, A. M., Crowder, C. M., Ousley, S. D., & Houck, M. M. (2014). Error and its meaning in forensic science. *Journal of Forensic Sciences*, 59(1), 123–126. <https://doi.org/10.1111/1556-4029.12275>
- Ciardhuáin, S. Ó. (2004). An extended model of cybercrime investigations. *International Journal of Digital Evidence*, 3(1), 1–22.
- Cloppert, M. (2009). Is digital forensics a science? *SANS Cyber Security Blog*. <https://www.sans.org/blog/is-digital-forensics-a-science/>

- Cohen, F. (2013). *Digital forensic evidence examination* (5th ed.). Fred Cohen & Associates.
- Cohen, F., Lowrie, J., & Preston, C. (2011). The state of the science of digital evidence examination. In G. Peterson & S. Sheno (Eds.), *Advances in digital forensics VII, DigitalForensics 2011, IFIP Advances in information and communication technology, vol. 361* (pp. 3-21). Springer. https://doi.org/10.1007/978-3-642-24212-0_1
- Cole, S. A. (2001). *Suspect identities: A history of criminal identification and fingerprinting*. Harvard University Press.
- Collie, J. (2018). Digital forensic evidence—Flaws in the criminal justice system. *Forensic Science International*, 289, 154–155. <https://doi.org/10.1016/j.forsciint.2018.05.014>
- Collie, J., & Overill, R. (2020). DEEP: Extending the digital forensics process model for criminal investigations. *Athens Journal of Sciences*, 7(4), 225–240. <https://doi.org/10.30958/ajs.7-4-3>
- Collier, P. A., & Spaul, B. J. (1992). A forensic methodology for countering computer crime. *Artificial Intelligence Review*, 6, 203–215. <https://doi.org/10.1007/BF00150234>
- Collins, H. M. (1981). Stages in the empirical programme of relativism. *Social Studies of Science*, 11(1), 3–10. <https://doi.org/10.1177/030631278101100101>
- Computer History Museum. (2021). *The timeline of computer history*. Computer History Museum. <https://www.computerhistory.org/timeline/>
- Cook, R., Evett, I. W., Jackson, G., Jones, P. J., & Lambert, J. A. (1998a). A hierarchy of propositions: Deciding which level to address in casework. *Science & Justice*, 38(4), 231–239. [https://doi.org/10.1016/s1355-0306\(98\)72117-3](https://doi.org/10.1016/s1355-0306(98)72117-3)
- Cook, R., Evett, I. W., Jackson, G., Jones, P. J., & Lambert, J. A. (1998b). A model for case assessment and interpretation. *Science & Justice*, 38(3), 151–156. [https://doi.org/10.1016/s1355-0306\(98\)72099-4](https://doi.org/10.1016/s1355-0306(98)72099-4)
- Cook, T. (2016). *Blackstone's senior investigating officers' handbook* (4th ed.). Oxford University Press.
- Cooper, G. S., & Meterko, V. (2019). Cognitive bias research in forensic science: A systematic review. *Forensic Science International*, 297, 35–46. <https://doi.org/10.1016/j.forsciint.2019.01.016>
- Costa, S., & Santos, F. (2019). The social life of forensic evidence and the epistemic sub-cultures in an inquisitorial justice system: Analysis of Saltão case. *Science & Justice*, 59(5), 471–479. <https://doi.org/10.1016/j.scijus.2019.06.003>

- Costantini, S., De Gasperis, G., & Olivieri, R. (2019). Digital forensics and investigations meet artificial intelligence. *Annals of Mathematics and Artificial Intelligence*, 86, 193–229. <https://doi.org/10.1007/s10472-019-09632-y>
- Creswell, J. W. (2013). *Qualitative inquiry and research design. Choosing among five approaches* (3rd ed.). Sage.
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage.
- Dahl, J. Y. (2008). Another side of the story: Defence lawyers' views on DNA evidence. In K. F. Aas, H. O. I. Gundhus, & H. M. Lomell (Eds.), *Technologies of insecurity* (pp. 233–251). Routledge-Cavendish.
- Dahl, J. Y. (2009). *DNA-det sikreste av det sikre eller...?: En sosiologisk studie av usikkerheter knyttet til bruk av DNA i strafferettspleien*. PhD thesis, Norwegian University of Science and Technology. NTNU Open. <http://hdl.handle.net/11250/268114>
- Dahl, J. Y., & Sætnan, A. R. (2009). “It all happened so slowly”– On controlling function creep in forensic DNA databases. *International Journal of Law, Crime and Justice*, 37(3), 83–103. <https://doi.org/10.1016/j.ijlcj.2009.04.002>
- Daston, L. (1992). Objectivity and the escape from perspective. *Social Studies of Science*, 22(4), 597–618. <https://doi.org/10.1177/030631292022004002>
- Daston, L., & Galison, P. (2007). *Objectivity*. Zone Books.
- Dean, G. (2000). *The experience of investigation for detectives*. PhD thesis, Queensland University of Technology.
- Denzin, N. K. (1970). *The research act in sociology*. Aldine.
- Doherty, N. F., Coombs, C. R., & Loan-Clarke, J. (2006). A re-conceptualization of the interpretive flexibility of information technologies: Redressing the balance between the social and the technical. *European Journal of Information Systems*, 15(6), 569–582. <https://doi.org/10.1057/palgrave.ejis.3000653>
- Dror, I. E. (2016). A hierarchy of expert performance. *Journal of Applied Research in Memory and Cognition*, 5(2), 121–127. <https://doi.org/10.1016/j.jarmac.2016.03.001>
- Dror, I. E. (2017). Human expert performance in forensic decision making: Seven different sources of bias. *Australian Journal of Forensic Sciences*, 49(5), 541–547. <https://doi.org/10.1080/00450618.2017.1281348>

- Dror, I. E. (2020). Cognitive and human factors in expert decision making: Six fallacies and the eight sources of bias. *Analytical Chemistry*, 92(12), 7998–8004.
<https://doi.org/10.1021/acs.analchem.0c00704>
- Dror, I. E., & Kukucka, J. (2021). Linear Sequential Unmasking–Expanded (LSU-E): A general approach for improving decision making as well as minimizing noise and bias. *Forensic Science International: Synergy*, 3, 100161.
<https://doi.org/10.1016/j.fsisyn.2021.100161>
- Dror, I. E., & Mnookin, J. L. (2010). The use of technology in human expert domains: Challenges and risks arising from the use of automated fingerprint identification systems in forensic science. *Law, Probability and Risk*, 9(1), 47–67.
<https://doi.org/10.1093/lpr/mgp031>
- Dror, I. E., Thompson, W. C., Meissner, C. A., Kornfield, I., Krane, D., Saks, M., & Risinger, M. (2015). Letter to the Editor-Context management toolbox: A Linear Sequential Unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *Journal of Forensic Sciences*, 60(4), 1111-1112.
<https://doi.org/10.1111/1556-4029.12805>
- Duarte, D. E. (2021). The making of crime predictions: Sociotechnical assemblages and the controversies of governing future crime. *Surveillance & Society*, 19(2), 199–215.
<https://doi.org/10.24908/ss.v19i2.14261>
- Earwaker, H., Nakhaeizadeh, S., Smit, N. M., & Morgan, R. M. (2020). A cultural change to enable improved decision-making in forensic science: A six phased approach. *Science & Justice*, 60(1), 9–19. <https://doi.org/10.1016/j.scijus.2019.08.006>
- Ekfeldt, J. (2016). *Om informationstekniskt bevis*. PhD thesis, Stockholm University. DiVA.
<http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Asu%3Adiva-125286>
- ENFSI. (2015a). *Best practice manual for the forensic examination of digital technology, ENFSI-BPM-FIT-01, Version 01 (November 2015)*. European Network of Forensic Science Institutes. https://enfsi.eu/wp-content/uploads/2016/09/1._forensic_examination_of_digital_technology_0.pdf
- ENFSI. (2015b). *ENFSI guideline for evaluative reporting in forensic science. Strengthening the evaluation of forensic results across Europe (STEOFRAE)*. European Network of Forensic Science Institutes. https://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf
- ENFSI. (2022). *History*. European Network of Forensic Science Institutes.
<https://enfsi.eu/history/>

- Erlandsen, T. E. (2019). *Fallacies when evaluating digital evidence among prosecutors in the Norwegian police service*. Master's thesis, The Norwegian University of Science and Technology. NTNU Open. <http://hdl.handle.net/11250/2617771>
- Evelt, I. W., Jackson, G., & Lambert, J. A. (2000). More on the hierarchy of propositions: Exploring the distinction between explanations and propositions. *Science & Justice*, 40(1), 3–10. [https://doi.org/10.1016/s1355-0306\(00\)71926-5](https://doi.org/10.1016/s1355-0306(00)71926-5)
- Fahsing, I. A. (2016). *The making of an expert detective. Thinking and deciding in criminal investigations*. PhD thesis, University of Gothenburg. PIA. <http://hdl.handle.net/11250/2428006>
- Farmer, D., & Venema, W. (2005). *Forensic discovery* (1st ed.). Addison Wesley Professional.
- Federal Bureau of Investigation. (1997). *History of the FBI laboratory (April 15, 1997)*. <https://irp.fas.org/agency/doj/oig/fbilab1/labpr.htm>
- Ferguson, R. I., Renaud, K., Wilford, S., & Irons, A. (2020). PRECEPT: A framework for ethical digital forensics investigations. *Journal of Intellectual Capital*, 21(2), 257–290. <https://doi.org/10.1108/JIC-05-2019-0097>
- Flaglien, A. O. (2018). The digital forensics process. In A. Årnes (Ed.), *Digital forensics* (pp. 13–49). Wiley.
- Flinders, K. (2021, 6 April). Demands for changes to ‘barmy’ rules on digital evidence have government’s ear. *Computerweekly.com*. <https://www.computerweekly.com/news/252498901/Demands-for-changes-to-barmy-rules-on-digital-evidence-have-governments-ear>
- Floridi, L. (1999). *Philosophy and computing: An introduction*. Routledge.
- Forensic Science Regulator. (2020). *Guidance. Cognitive bias effects relevant to forensic science examinations (FSR-G-217, Issue 2)*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/914259/217_FSR-G-217_Cognitive_bias_appendix_Issue_2.pdf
- Friheim, I. (2016). *Practical use of dual tool verification in computer forensics*. Master's thesis, University College Dublin.
- Gabbert, F., Hope, L., & Confrey, M. (2018). Witness testimony. In A. Griffiths & R. Milne (Eds.), *The psychology of criminal investigation* (pp. 113-132). Routledge.
- Gardner, B. O., Kelley, S., Murrie, D. C., & Dror, I. E. (2019). What do forensic analysts consider relevant to their decision making? *Science & Justice*, 59(5), 516–523. <https://doi.org/10.1016/j.scijus.2019.04.005>

- Garfinkel, S. (2012). Lessons learned writing digital forensics tools and managing a 30TB digital evidence corpus. *Digital Investigation*, 9, Supplement, 80–89.
<https://doi.org/10.1016/j.diin.2012.05.002>
- Garfinkel, S., Farrell, P., Roussev, V., & Dinolt, G. (2009). Bringing science to digital forensics with standardized forensic corpora. *Digital Investigation*, 6, Supplement, 2–11. <https://doi.org/10.1016/j.diin.2009.06.016>
- Garfinkel, S. L. (2010). Digital forensics research: The next 10 years. *Digital Investigation*, 7, Supplement, 64–73. <https://doi.org/10.1016/j.diin.2010.05.009>
- Garrett, B. L. (2021). *Autopsy of a crime lab*. University of California Press.
- Geradts, Z. (2011). ENFSI forensic IT working group. *Digital Investigation*, 8(2), 94–95.
<https://doi.org/10.1016/j.diin.2011.09.003>
- Gladyshev, P. (2004). *Formalizing event reconstruction in digital investigations*. PhD thesis, University College Dublin.
- Glisson, W. B., Storer, T., & Buchanan-Wollaston, J. (2013). An empirical comparison of data recovered from mobile forensic toolkits. *Digital Investigation*, 10(1), 44–55.
<https://doi.org/10.1016/j.diin.2013.03.004>
- Gogolin, G. (2010). The digital crime tsunami. *Digital Investigation*, 7(1-2), 3–8.
<https://doi.org/10.1016/j.diin.2010.07.001>
- Granhag, P., & Ask, K. (2021). Psykologiska perspektiv på bevisvärdering. In P. Granhag, L. A. Strömvall, K. Ask & S. Landström (Eds.), *Handbok i rättspsykologi* (pp. 427–445). Liber.
- Griffiths, A., & Rachlew, A. (2018). From interrogation to investigative interviewing: The application of psychology. In A. Griffiths & R. Milne (Eds.), *The psychology of criminal investigation* (pp. 154–178). Routledge.
- Grigaliunas, S., Toldinas, J., Venckauskas, A., Morkevicius, N., & Damasevicius, R. (2021). Digital evidence object model for situation awareness and decision making in digital forensics investigation. *IEEE Intelligent Systems*, 36(5), 39–48.
<https://doi.org/10.1109/MIS.2020.3020008>
- Grobler, M. (2012). The need for digital evidence standardisation. *International Journal of Digital Crime and Forensics*, 4(2), 1–12. <https://doi.org/10.4018/jdcf.2012040101>
- Grut, S. (2020, 3 November). Feil i analyseverktøy gjør at politiet må gjennomgå minst 57 straffesaker. *NRK Beta*. <https://nrkbeta.no/2020/11/03/feil-i-analyseverktoy-gjor-at-politiet-ma-undersoke-flere-titalls-straffesaker-pa-nytt/>

- Gundhus, H. O. I. (2013). Experience or knowledge? Perspectives on new knowledge regimes and control of police professionalism. *Policing: A Journal of Policy and Practice*, 7(2), 178–194. <https://doi.org/10.1093/police/pas039>
- Gundhus, H. O. I., Talberg, N., & Wathne, C. T. (2022). From discretion to standardization: Digitalization of the police organization. *International Journal of Police Science & Management*, 24(1), 27–41. <https://doi.org/10.1177/14613557211036554>
- Guo, H., & Hou, J. (2018). Review of the accreditation of digital forensics in China. *Forensic Sciences Research*, 3(3), 194–201. <https://doi.org/10.1080/20961790.2018.1503526>
- Guttman, B., Laamanen, M. T., Russell, C., Atha, C., & Darnell, J. (2022). *Results from a black-box study for digital forensic examiners (NISTIR8412)*. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.IR.8412>
- Hald, C., & Rønn, K. V. (2013). Inledning. In C. Hald & K. V. Rønn (Eds.), *Om at opdage. Metodiske refleksjoner over politiets undersøgelsespraksis* (pp. 15-54). Samfundslitteratur.
- Hansen, H. A., Andersen, S., Axelsson, S., & Hopland, S. (2017). *Case study: A new method for investigating crimes against children*. Paper presented at the Conference on Digital Forensics, Security and Law, Florida, USA. <https://commons.erau.edu/cgi/viewcontent.cgi?article=1378&context=adfls>
- Haraldseid, S. (2021). *Kan du stikke opp og gå gjennom databeslaget? Framgangsmåter for innholdsanalyse av databeslag og behovet for metodisk støtte*. Master's thesis, The Norwegian Police University College. PIA. <https://hdl.handle.net/11250/2760497>
- Harrison, W. (2004). The digital detective: An introduction to digital forensics. *Advances in Computers*, 60, 75–119. [https://doi.org/10.1016/s0065-2458\(03\)60003-3](https://doi.org/10.1016/s0065-2458(03)60003-3)
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77–89. <https://doi.org/10.1080/19312450709336664>
- Heap, V., & Waters, J. (2019). *Mixed methods in criminology*. Routledge.
- Hewling, M. O. (2013). *Digital forensics: An integrated approach for the investigation of cyber/computer related crimes*. PhD thesis, University of Bedfordshire. UOBREP. <http://hdl.handle.net/10547/326231>
- Hestehave, N. K. (2021). *Coppers chasing usual suspects – An embedded search for a proactive performance and the detective métier*. PhD thesis, Aalborg University.
- Holmberg, L. (2014). Hva gjør politiet? In P. Larsson, H. O. I. Gundhus & R. Granér (Eds.), *Innføring i politivitenskap* (pp. 153-177). Cappelen Damm Akademisk.

- Holt, T. J., Bossler, A. M., & Seigfried-Spellar, K. C. (2017). *Cybercrime and digital forensics: An introduction* (2nd ed.). Routledge.
- Hon, G. (1995). Going wrong: To make a mistake, to fall into an error. *The Review of Metaphysics*, 49(1), 3–20. <http://www.jstor.org/stable/20129804>
- Horsman, G. (2018). Framework for Reliable Experimental Design (FRED): A research framework to ensure the dependable interpretation of digital data for digital forensics. *Computers & Security*, 73, 294–306. <https://doi.org/10.1016/j.cose.2017.11.009>
- Horsman, G. (2019). Formalising investigative decision making in digital forensics: Proposing the Digital Evidence Reporting and Decision Support (DERDS) framework. *Digital Investigation*, 28, 146–151. <https://doi.org/10.1016/j.diin.2019.01.007>
- Horsman, G. (2020). Opinion: Does the field of digital forensics have a consistency problem? *Forensic Science International: Digital Investigation*, 33, 300970. <https://doi.org/10.1016/j.fsidi.2020.300970>
- Horsman, G. (2021). The different types of reports produced in digital forensic investigations. *Science & Justice*, 61(5), 627–634. <https://doi.org/10.1016/j.scijus.2021.06.009>
- Horsman, G., & Sunde, N. (2020). Part 1: The need for peer review in digital forensics. *Forensic Science International: Digital Investigation*, 35, 301062. <https://doi.org/10.1016/j.fsidi.2020.301062>
- Horsman, G., & Sunde, N. (2022). Unboxing the digital forensic investigation process. *Science & Justice*, 62(2), 171–180. <https://doi.org/10.1016/j.scijus.2022.01.002>
- Houck, M. M. (2009). Trace evidence. In J. Fraser & R. Williams (Eds.), *Handbook of forensic science* (pp. 166–195). Willan Publishing.
- Humphries, G., Nordvik, R., Manifavas, H., Cobley, P., & Sorell, M. (2021). Law enforcement educational challenges for mobile forensics. *Forensic Science International: Digital Investigation*, 38, Supplement, 301129. <https://doi.org/10.1016/j.fsidi.2021.301129>
- Humphries, G. L. (2019). *Educating the effective digital forensics practitioner: Academic, professional, graduate and student perspectives*. PhD thesis, Canterbury Christ Church University.
- Innes, M. (2003). *Investigating murder: Detective work and the police response to criminal homicide*. Oxford University Press.
- Innes, M., Brookman, F., & Jones, H. (2021). “Mosaicking”: Cross construction, sense-making and methods of police investigation. *Policing: An International Journal*, 44(4), 708–721. <https://doi.org/10.1108/pijpsm-02-2021-0028>

- INTERPOL. (2019). *Global guidelines for digital forensics laboratories*. INTERPOL Global Complex for Innovation.
- Jackson, G. (2011). *The development of case assessment and interpretation (CAI) in forensic science*. PhD thesis, University of Abertay Dundee.
https://rke.abertay.ac.uk/ws/portalfiles/portal/15382303/Jackson_2011_The_development_of_case_assessment_Redacted.pdf
- Jackson, G., Jones, S., Booth, G., Champod, C., & Evett, I. W. (2006). The nature of forensic science opinion - A possible framework to guide thinking and practice in investigation and in court proceedings. *Science & Justice*, 46(1), 33–44.
[https://doi.org/10.1016/s1355-0306\(06\)71565-9](https://doi.org/10.1016/s1355-0306(06)71565-9)
- Jahren, J. H. (2020). *Is the quality assurance in digital forensic work in the Norwegian police adequate?* Master's thesis, The Norwegian University of Science and Technology. NTNU Open. <https://hdl.handle.net/11250/2781174>
- Jakobsen, K. K. (2021). *Objektivitet og empati i avhør av fornærmede: En kvalitativ undersøkelse av norske politiavhør*. PhD thesis, University of Oslo. DUO.
<http://hdl.handle.net/10852/86483>
- James, J. I., & Gladyshev, P. (2013a). *Challenges with automation in digital forensic investigations*. arXiv. <https://doi.org/10.48550/arXiv.1303.4498>
- James, J. I., & Gladyshev, P. (2013b). A survey of digital forensic investigator decision processes and measurement of decisions based on enhanced preview. *Digital Investigation*, 10(2), 148–157. <https://doi.org/10.1016/j.diin.2013.04.005>
- James, S. H., & Nordby, J. J. (2002). *Forensic science: An introduction to scientific and investigative techniques*. CRC Press.
- Jaquet-Chiffelle, D.-O., & Casey, E. (2021). A formalized model of the trace. *Forensic Science International*, 327, 110941. <https://doi.org/10.1016/j.forsciint.2021.110941>
- Jones, H., Brookman, F., Williams, R., & Fraser, J. (2020). We need to talk about dialogue: Accomplishing collaborative sensemaking in homicide investigations. *The Police Journal*, 94(4), 572–589. <https://doi.org/10.1177/0032258x20970999>
- Jonsson, K., & Holmström, J. (2005). Ubiquitous computing and the double immutability of remote diagnostics technology: An exploration into six cases of remote diagnostics technology use. In C. Sørensen, Y. Yoo, K. Lyytinen, & J. I. DeGross (Eds.), *Designing ubiquitous information environments: Socio-technical issues and challenges*. IFIP — The International Federation for Information Processing, vol 185 (pp. 153–167). Springer. https://doi.org/10.1007/0-387-28918-6_13

- Jordaan, J. (2021). *SANS Digital forensics survey: Digital forensic essentials and why foundations matter*. <https://www.sans.org/account/login?url=/white-papers/40420>
- Julian, R., Howes, L., & White, R. (2021). *Critical forensic studies*. Routledge.
- Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Brown Spark Little.
- Karie, N. M., Kebande, V. R., Venter, H. S., & Choo, K.-K. R. (2019). On the importance of standardising the process of generating digital forensic reports. *Forensic Science International: Reports, 1*, 100008. <https://doi.org/10.1016/j.fsir.2019.100008>
- Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition, 2*(1), 42–52. <https://doi.org/10.1016/j.jarmac.2013.01.001>
- Kaufmann, M. (2017). The co-construction of crime predictions: Dynamics between digital data, software and human beings. In N. R. Fyfe, H. O. I. Gundhus & K. V. Rønn (Eds.), *Moral issues in intelligence led policing* (pp. 143-160). Routledge.
- Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *The British Journal of Criminology, 59*(3), 674–692. <https://doi.org/10.1093/bjc/azy060>
- Kelly, L., Sachan, S., Ni, L., Almaghrabi, F., Allmendinger, R., & Chen, Y.-W. (2020). Explainable artificial intelligence for digital forensics: Opportunities, challenges and a drug testing case study. In B. S. Shetti & P. Shetty (Eds.), *Digital forensic science* (pp. 1-19). IntechOpen. <https://doi.org/10.5772/intechopen.93310>
- Kimmelman, J. (2021). Clinical Trials to authors: Please pre-register your studies! *Clinical Studies, 18*(6), 645–646. <https://doi.org/10.1177/17407745211057186>
- King, G. L. (2006). *Forensics plan guide* (Global information assurance certification paper). SANS Institute.
- Kirk, P. L. (1953). *Crime investigation: Physical evidence and the police laboratory*. Interscience Publishers.
- Kloess, J. A., Woodhams, J., & Hamilton-Giachritsis, C. E. (2021). The challenges of identifying and classifying child sexual exploitation material: Moving towards a more ecologically valid pilot study with digital forensics analysts. *Child Abuse & Neglect, 118*, 105166. <https://doi.org/10.1016/j.chiabu.2021.105166>
- Knorr-Cetina, K. D. (1981). *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Pergamon Press.

- Krippendorff, K. (2011). *Computing Krippendorff's Alpha-Reliability*. Unpublished manuscript.
- Kruse, C. (2016). *The social life of forensic evidence*. University of California Press.
- Kukucka, J. & Dror, I. E. (2022). Human factors in forensic science: Psychological causes of bias and error. In D. DeMatteo & K. Scherr (Eds.), *The Oxford Handbook of Psychology and Law* (forthcoming). Oxford University Press.
- Larsson, P., Gundhus, H. O. I., & Granér, R. (2014). Politivitenskap - en introduksjon. In P. Larsson, H. O. I. Gundhus, & R. Granér (Eds.), *Innføring i politivitenskap* (pp. 15–28). Cappelen Damm Akademisk.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.
- Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In W. E. Bijker & J. Law (Eds.), *Shaping technology/building society: Studies in sociotechnical change* (Vol. 1, pp. 225-258). MIT Press.
- Latour, B. (2005). *Reassembling the social. An introduction to Actor-Network-Theory*. Oxford University Press.
- Latour, B., & Woolgar, S. (1979/1986). *Laboratory life: The construction of scientific facts* (2nd ed. with a new subtitle). Princeton University Press.
- Lawless, C. (2022). *Forensic science. A sociological introduction* (2nd ed.). Routledge.
- Leedy, P. D., & Ormrod, J. E. (2014). *Practical research planning and design*. Pearson Education Limited.
- Leese, M. (2021). Security as socio-technical practice: Predictive policing and (non-) automation. *Swiss Political Science Review*, 27(1), 150–157.
<https://doi.org/10.1111/spsr.12432>
- Lempert, R. O., Gross, S. O., & Liebman, J. S. (2000). *Modern approach to evidence* (3rd ed.). West Group.
- Lentz, L. W., & Sunde, N. (2020). The use of historical call data records as evidence in the criminal justice system - Lessons learned from the Danish telecom scandal. *Digital Evidence and Electronic Signature Law Review*, 18, 1–17.
<https://doi.org/10.14296/deeslr.v18i0.5235>
- Leppänen, A., & Kankaanranta, T. (2017). Cybercrime investigation in Finland. *Journal of Scandinavian Studies in Criminology and Crime Prevention*, 18(2), 157–175.
<https://doi.org/10.1080/14043858.2017.1385231>

- Lillis, D., Becker, B., O'Sullivan, T., & Scanlon, M. (2016). *Current challenges and future research areas for digital forensic investigation*. Paper presented at the 11th ADFSLS Conference on Digital Forensics, Security and Law (CDFSL 2016), Daytona Beach, Florida, USA.
- Lipton, P. (1991). *Inference to the best explanation*. Routledge.
- Lomell, H. M. (2004). Targeting the unwanted: Video surveillance and categorical exclusion in Oslo, Norway. *Surveillance & Society*, 2(2/3).
<https://doi.org/10.24908/ss.v2i2/3.3382>
- Lopez, E. M., Moon, S. Y., & Park, J. H. (2016). Scenario-based digital forensics challenges in cloud computing. *Symmetry*, 8(10), 2–20. <https://doi.org/10.3390/sym8100107>
- Losavio, M., Seigfried-Spellar, K. C., & Sloan, J. J. (2016). Why digital forensics is not a profession and how it can become one. *Criminal Justice Studies*, 29(2), 143–162.
<https://doi.org/10.1080/1478601x.2016.1170281>
- Luciano, L., Baggili, I., Topor, M., Casey, P., & Breitingner, F. (2018). *Digital forensics in the next five years*. Paper presented at the 13th International Conference on Availability, Reliability and Security (ARES 2018), Hamburg, Germany.
<https://doi.org/10.1145/3230833.3232813>
- Lundgaard, J. (2019). *Kritisk kunnskap. Meningsdannelse og beslutningsprosesser ved politiets operasjonssentral*. PhD thesis, University of Oslo.
- Lyle, J. R., Guttman, B., Butler, J. M., Sauerwein, K., Reed, C., & Lloyd, C. E. (2022). *Digital investigation techniques: A NIST scientific foundation review (NISTIR 8354-DRAFT)*. National Institute of Standards and Technology.
<https://nvlpubs.nist.gov/nistpubs/ir/2022/NIST.IR.8354-draft.pdf>
- Maguire, M. (2008). Criminal investigation and crime control, 2nd ed. In T. Newburn (Ed.), *Handbook of policing* (Vol. 2, pp. 430-464). Routledge.
- Manoogian, J., & Benson, B. (2018). *Cognitive bias codex*.
https://commons.wikimedia.org/wiki/File:Cognitive_bias_codex_en.svg
- Marciniak, D. (2021). *Data-driven policing: How digital technologies transform the practice and governance of policing*. PhD thesis, University of Essex.
- Marshall, A. M., & Paige, R. (2018). Requirements in digital forensics method definition: Observations from a UK study. *Digital Investigation*, 27, 23–29.
<https://doi.org/10.1016/j.diin.2018.09.004>
- Marsico, C. V. (2005). *Digital music device forensics*. Master's thesis, Purdue University. CERIAS. https://www.cerias.purdue.edu/assets/pdf/bibtex_archive/2005-27.pdf

- Martire, K. A., Kemp, R. I., Sayle, M., & Newell, B. R. (2014). On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Science International*, *240*, 61–68.
<https://doi.org/10.1016/j.forsciint.2014.04.005>
- Marx, G. T., & Guzik, K. (2017). The uncertainty principle: Qualification, contingency, and fluidity in technology and social control. In M. R. McGuire & T. J. Holt (Eds.), *The Routledge handbook of technology, crime and justice* (pp. 481–502). Routledge.
- McCartney, C. (2019). Streamlined forensic reporting: Rhetoric and reality. *Forensic Science International: Synergy*, *1*, 83–85. <https://doi.org/10.1016/j.fsisyn.2019.04.004>
- McKemmish, R. (1999). *What is forensic computing?* Australian Institute of Criminology.
- McKemmish, R. (2008). When is digital evidence forensically sound? In I. Ray & S. Sheinoi (Eds.), *Advances in digital forensics IV, DigitalForensics 2008, IFIP – The International Federation for Information Processing*, vol. 285 (pp. 3–15). Springer.
https://doi.org/10.1007/978-0-387-84927-0_1
- Meterko, V., & Cooper, G. (2021). Cognitive biases in criminal case evaluation: A review of the research. *Journal of Police and Criminal Psychology*, *37*, 101–122.
<https://doi.org/10.1007/s11896-020-09425-8>
- Meyers, M., & Rogers, M. (2004). Computer forensics: The need for standardization and certification. *International Journal of Digital Evidence*, *3*(2), 1–11.
- Montasari, R., Peltola, P., & Evans, D. (2015). Integrated Computer Forensics Investigation Process Model (ICFIPM) for computer crime investigations. In H. Jahankhani, A. Carlile, B. Akhgar, A. Taal, A. Hessami & A. Hosseinian-Far, (Eds.), *Global security, safety and sustainability: Tomorrow's challenges of cyber security. ICGS3 2015. Communications in computer and information science*, vol. 534 (pp. 83-95). Springer.
https://doi.org/10.1007/978-3-319-23276-8_8
- Muir, R., & Walcott, S. (2021). *Unleashing the value of digital forensics*.
https://www.police-foundation.org.uk/2017/wp-content/uploads/2010/10/value_of_digital_forensics.pdf
- Narwal, B., & Goel, N. (2020). A walkthrough of digital forensics and its tools. *TEST Engineering & Management*, *82*, 13757–13764.
- National Institute of Justice. (2008). *Electronic crime scene investigation: A guide for first responders* (2nd ed.). U.S. Department of Justice.
<https://www.ojp.gov/pdffiles1/nij/219941.pdf>

- National Police Chiefs' Council. (2020). *Digital forensics science strategy*. National Police Chiefs' Council.
<https://www.npcc.police.uk/FreedomofInformation/Reportsreviewsandresponsestoconsultations.aspx>
- National Research Council. (2009). *Strengthening forensic science in the United States: A path forward*. National Academic Press.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nili, A., Tate, M., & Barros, A. (2017). *A critical analysis of inter-coder reliability methods in information systems research*. Paper presented at the 28th Australasian Conference on Information Systems, University of Tasmania, Australia.
- Noblett, M. G., Pollitt, M., & Presley, L. A. (2000). Recovering and examining computer forensic evidence. *Forensic Science Communications*, 2(4).
- Nordvik, R., Stoykova, R., Franke, K., Axelsson, S., & Toolan, F. (2021). Reliability validation for file system interpretation. *Forensic Science International: Digital Investigation*, 37, 301174. <https://doi.org/10.1016/j.fsidi.2021.301174>
- Nygaard, L. P., & Solli, K. (2020). *Strategies for writing a thesis by publication in the social sciences and humanities*. Routledge.
- Oliveira Jr, E., Zorzo, A. F., & Neu, C. V. (2020). Towards a conceptual model for promoting digital forensics experiments. *Forensic Science International: Digital Investigation*, 35, 301014. <https://doi.org/10.1016/j.fsidi.2020.301014>
- Olivier, M., & Gruner, S. (2013). On the scientific maturity of digital forensics research. In G. Peterson & S. Sheno (Eds.), *Advances in digital forensics IX, DigitalForensics 2013, IFIP Advances in information and communication technology, vol. 410* (pp. 33-49). Springer. https://doi.org/10.1007/978-3-642-41148-9_3
- Olivier, M. S. (2016a). Digital forensic science: A manifesto. *South African Computer Journal*, 28(2), 46–59. <https://doi.org/10.18489/sacj.v28i2.442>
- Olivier, M. (2016). On a scientific theory of digital forensics. In G. Peterson & S. Sheno (Eds.), *Advances in digital forensics XII, DigitalForensics 2016, IFIP Advances in information and communication technology, vol. 484* (pp. 3-24). Springer. https://doi.org/10.1007/978-3-319-46279-0_1
- Osborne, G., Thinyane, H., & Slay, J. (2012). Visualizing information in digital forensics. In G. Peterson & S. Sheno (Eds.), *Advances in digital forensics VIII, DigitalForensics*

- 2012, *IFIP Advances in information and communication technology*, vol. 383 (pp. 35-47). Springer. https://doi.org/10.1007/978-3-642-33962-2_3
- Overill, R., & Chow, K-P. (2018). Measuring evidential weight in digital forensic investigations. In G. Peterson & S. Sheno (Eds.), *Advances in digital forensics XIV, DigitalForensics 2018, IFIP Advances in information and communication technology*, vol. 532 (pp. 3-10). Springer. https://doi.org/10.1007/978-3-319-99277-8_1
- Overill, R. E., & Collie, J. (2021). Quantitative evaluation of the results of digital forensic investigations: A review of progress. *Forensic Sciences Research*, 6(1), 13–18. <https://doi.org/10.1080/20961790.2020.1837429>
- Page, H., Horsman, G., Sarna, A., & Foster, J. (2018). A review of quality procedures in the UK forensic sciences: What can the field of digital forensics learn? *Science & Justice*, 59(1), 83–92. <https://doi.org/10.1016/j.scijus.2018.09.006>
- Palmer, G. (2001). *A road map for digital forensic research*. Technical report (TDR-T001-01) for Digital Forensic Research Workshop (DFRWS), New York.
- Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, 34(5 Pt 2), 1189–1208. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1089059/>
- Pennington, N., & Hastie, R. (1993). *The story model for juror decision making*. Cambridge University Press.
- Pohl, R. F. (2022). What are cognitive illusions? In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgement, and memory* (3rd ed.) (pp. 3-23). Routledge.
- Politidirektoratet. (2017). *Retningslinjer for bruk av etterforskningsplan. Versjon 2.0*. Politidirektoratet.
- Pollitt, M. (1995). Computer forensics: An approach to evidence in cyberspace. *Proceedings of the 18th National Information Systems Security Conference, Maryland, USA*. <https://csrc.nist.gov/CSRC/media/Publications/conference-paper/1995/10/10/proceedings-of-the-18th-nissc-1995/documents/1995-18th-NISSC-proceedings-vol-1.pdf>
- Pollitt, M. (2010). *A history of digital forensics*. Paper presented at the IFIP Advances in Information and Communication Technology.
- Pollitt, M., Casey, E., Jaquet-Chiffelle, D., & Gladyshev, P. (2018). *A framework for harmonizing forensic science practices and digital/multimedia evidence*. The

- Organization of Scientific Area Committees for Forensic Science (OSAC).
https://www.nist.gov/system/files/documents/2018/01/10/osac_ts_0002.pdf
- Powell, A., Stratton, G., & Cameron, R. (2018). *Digital criminology: Crime and justice in digital society*. Routledge.
- President's Council of Advisors on Science & Technology. (2016). *Report to the President, forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*. Executive Office of the President of the United States, President's Council of Advisors on Science & Technology.
- Quick, D., & Choo, K.-K. R. (2014). Impacts of increasing volume of digital forensic data: A survey and future research challenges. *Digital Investigation*, 11(4), 273–294.
<https://doi.org/10.1016/j.diin.2014.09.002>
- Rachlew, A. (2009). Justisfeil ved politiets etterforskning. Noen eksempler og forskningsbaserte tiltak, PhD thesis, University of Oslo. DUO Vitenarkiv.
<http://urn.nb.no/URN:NBN:no-23961>
- Raghavan, S. (2013). Digital forensic research: Current state of the art. *CSI Transactions on ICT*, 1(1), 91–114. <https://doi.org/10.1007/s40012-012-0008-7>
- Rappert, B., Wheat, H., & Wilson-Kovacs, D. (2021). Rationing bytes: Managing demand for digital forensic examinations. *Policing and Society*, 31(1), 52–65.
<https://doi.org/10.1080/10439463.2020.1788026>
- Reason, J. (1990). *Human error*. Cambridge University Press.
- Reedy, P. (2020). Interpol review of digital evidence 2016-2019. *Forensic Science International: Synergy*, 2, 489–520. <https://doi.org/10.1016/j.fsisyn.2020.01.015>
- Reiner, R. (2010). *The politics of the police*. Oxford University Press.
- Reith, M., Carr, C., & Gunsch, G. (2002). An examination of digital forensic models. *International Journal of Digital Evidence*, 1(3), 1–12.
- Risinger, D. M. (2013). Reservations about likelihood ratios (and some other aspects of forensic ‘Bayesianism’). *Law, Probability and Risk*, 12(1), 63–73.
<https://doi.org/10.1093/lpr/mgs011>
- Rogers, M. (2017). Technology and digital forensics. In M. R. McGuire & T. J. Holt (Eds.), *The Routledge handbook of technology, crime and justice* (pp. 406–416). Routledge.
- Runhovde, S. R. (2017). *Policing the illegal trade in wildlife. A study of Norway and Uganda*. PhD thesis, University of Oslo.

- Ryser, E., Spichiger, H., & Casey, E. (2020). Structured decision making in investigations involving digital and multimedia evidence. *Forensic Science International: Digital Investigation*, 34, 301015. <https://doi.org/10.1016/j.fsidi.2020.301015>
- Rønn, K. V. (2013). Mistanke - Hypoteser og forklaringer i opdagelsesarbejdet. In C. Hald & K. V. Rønn (Eds.), *Om at opdage. Metodiske refleksjoner over politiets undersøgelsespraksis* (pp. 255–299). Samfundslitteratur.
- Rønn, K. V. (2022). The multifaceted norm of objectivity in intelligence practices. *Intelligence and National Security*, 1-15. <https://doi.org/10.1080/02684527.2022.2076331>
- Sanders, C. (2021). *The analyst mindset: A cognitive skills assessment of digital forensic analysts*. PhD Thesis, Baylor University.
- Santos, F. (2014). Making sense of the story—The dialogues between the police and forensic laboratories in the construction of DNA evidence. *New Genetics and Society*, 33(2), 181–203. <https://doi.org/10.1080/14636778.2014.916186>
- Schjolberg, S. (2019). *The history of cybercrime* (3rd ed. (Vol. 13)). Nordstedt: Cybercrime Research Institute.
- Schniederjans, D., & Schniederjans, M. (2015). Quality management and innovation: New insights on a structural contingency framework. *International Journal of Quality Innovation*, 1(1), 1–20. <https://doi.org/10.1186/s40887-015-0004-8>
- Searston, R. A., Thompson, M. B., Robson, S. G., Corbett, B. J., Ribeiro, G., Edmond, G., & Tangen, J. M. (2019). Truth and transparency in expertise research. *Journal of Expertise/December*, 2(4), 199-209. https://www.journalofexpertise.org/articles/volume2_issue4/JoE_2_4_Searston.html
- Shadish, W., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth, Cengage Learning.
- Sharma, P., Arora, D., & Sakthivel, T. (2020). Enhanced forensic process for improving mobile cloud traceability in cloud-based mobile applications. *Procedia Computer Science*, 167, 907–917. <https://doi.org/10.1016/j.procs.2020.03.390>
- Shaw, M. (2001). *The coming-of-age of software architecture research*. Paper presented at the 23rd International Conference on Software Engineering. ICSE 2001, Toronto, Canada.
- Silvast, A., & Foulds, C. (2022). *Sociology of interdisciplinarity. The dynamics of energy research*. Palgrave Macmillan.
- Skjølvold, T. M. (2015). *Vitenskap, teknologi og samfunn: En introduksjon til STS*. Cappelen Damm Akademisk.

- Skre, A. B. (2020). Investigation plans as a tool for managing investigations in Norway. In X. Agirre, M. Bergsmo, S. De Smet, & C. Stahn (Eds.), *Quality control in criminal investigation* (pp. 887–902). Torkel Opsahl Academic EPublisher.
- Sommer, P. (2010). Forensic science standards in fast-changing environments. *Science & Justice*, 50(1), 12–17. <https://doi.org/10.1016/j.scijus.2009.11.006>
- Sorensen, M. S. (2019). Flaws in cellphone evidence prompt review of 10,000 verdicts in Denmark. *The New York Times*.
<https://www.nytimes.com/2019/08/20/world/europe/denmark-cellphone-data-courts.html>
- Spellman, B. A., Eldridge, H., & Bieber, P. (2021). Challenges to reasoning in forensic science decisions. *Forensic Science International: Synergy*, 4, 100200.
<https://doi.org/10.1016/j.fsisyn.2021.100200>
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3), 387–420.
<https://doi.org/10.1177/030631289019003001>
- Stelfox, P. (2009). *Criminal investigation: An introduction to principles and practice*. Routledge.
- Stoykova, R. (2021). Digital evidence: Unaddressed threats to fairness and the presumption of innocence. *Computer Law & Security Review*, 42, 105575.
<https://doi.org/10.1016/j.clsr.2021.105575>
- Stoykova, R., Andersen, S., Franke, K., & Axelsson, S. (2022). Reliability assessment of digital forensic investigations in the Norwegian police. *Forensic Science International: Digital Investigation*, 40, 301351.
<https://doi.org/10.1016/j.fsidi.2022.301351>
- Stratton, G., Powell, A., & Cameron, R. (2017). Crime and justice in digital society: Towards a 'digital criminology'? *International Journal for Crime, Justice and Social Democracy*, 6(2), 17–33. <https://doi.org/10.5204/ijcjsd.v6i2.355>
- Sunde, N. (2017). *Non-technical sources of errors when handling digital evidence within a criminal investigation*. Master's thesis, Norwegian University of Science and Technology. NTNU Open. <http://hdl.handle.net/11250/2472259>
- Sunde, N. (2020a). Digitale bevis i norske gjenåpningsbegjæringer – kan vi utelukke systematiske feil? *Tidsskrift for strafferett*, 20(1), 18-37.
<https://doi.org/10.18261/issn.0809-9537-2020-01-03>.

- Sunde, N. (2020b). Structured hypothesis development in criminal investigation: A method aimed at providing a broad and objective starting point for a criminal investigation. *The Police Journal*, 95(2), 276-295. <https://doi.org/10.1177/0032258X20982328>
- Sunde, N. (2022a). Min smartmobil er min borg. *Tidsskrift for strafferett*, 22(1), 25-46. <https://doi.org/10.18261/strafferett.22.1.2>
- Sunde, N. (2022b). Process modelling in digital forensics. (Unpublished manuscript).
- Sunde, N., & Horsman, G. (2021). Part 2: The Phase-oriented Advice and Review Structure (PARS) for digital forensic investigations. *Forensic Science International: Digital Investigation*, 36, 301074. <https://doi.org/10.1016/j.fsidi.2020.301074>
- SWGDE. (2016). *SWGDE Digital & multimedia evidence glossary. Version: 3.0 (June 23, 2016)*. Scientific Working Group on Digital Evidence. <https://drive.google.com/file/d/1ZZwOqgVOWo6qDeoJqv6VKafY2i1RJI2B/view>
- SWGDE. (2018). *SWGDE Establishing confidence in digital and multimedia evidence forensic results by error mitigation analysis. Version: 2.0 (Nov. 20, 2018)*. Scientific Working Group on Digital Evidence. https://drive.google.com/file/d/1pK_6eveU8Wb9TC9DvVpw1XNKPndwokKk/view
- Tart, M. (2020). Opinion evidence in cell site analysis. *Science & Justice*, 60(4), 363–374. <https://doi.org/10.1016/j.scijus.2020.02.002>
- Tecuci, G., Schum, D. A., Marcu, D., & Boicu, M. (2016). *Intelligence analysis as discovery of evidence, hypotheses, and arguments: Connecting the dots*. Cambridge University Press.
- Tewelde, S., Gruner, S., & Olivier, M. S. (2015). Notions of hypothesis in digital forensics. In G. Peterson, & S. Sheno (Eds.), *Advances in digital forensics XI, DigitalForensics 2015, IFIP Advances in information and communication technology, vol. 462* (pp. 29-43). Springer. https://dx.doi.org/10.1007/978-3-319-24123-4_2
- The Law Commission. (1997). Evidence in criminal proceedings: hearsay and related topics. https://www.lawcom.gov.uk/app/uploads/2015/03/lc245_Legislating_the_Criminal_Code_Evidence_in_Criminal_Proceedings.pdf
- Thompson, P., & Manning, M. (2021). Missed opportunities in digital investigation. In H. Jahankhani, A. Jamal & S. Lawson (Eds.), *Cybersecurity, privacy and freedom protection in the connected world* (pp. 101–122). Springer.
- Thompson, W. C., & Newman, E. J. (2015). Lay understanding of forensic statistics: Evaluation of random match probabilities, likelihood ratios, and verbal equivalents. *Law and Human Behavior*, 39(4), 332–349. <https://doi.org/10.1037/lhb0000134>

- Tilstone, W., Hastrup, M. L., & Hald, C. (2013). *Fischer's techniques of crime scene investigation*. CRC Press.
- Tong, S., & Bowling, B. (2006). Art, craft and science of detective work. *The Police Journal*, 79(4), 323–329. <https://doi.org/10.1350/pojo.2006.79.4.323>
- Tully, G., Cohen, N., Compton, D., Davies, G., Isbell, R., & Watson, T. (2020). Quality standards for digital forensics: Learning from experience in England & Wales. *Forensic Science International: Digital Investigation*, 32, 200905. <https://doi.org/10.1016/j.fsidi.2020.200905>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, New Series*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- van Beek, H. M. A., van den Bos, J., Boztas, A., van Eijk, E. J., Schramp, R., & Ugen, M. (2020). Digital forensics as a service: Stepping up the game. *Forensic Science International: Digital Investigation*, 35, 301021. <https://doi.org/10.1016/j.fsidi.2020.301021>
- Van Buskirk, E., & Liu, V. T. (2006). Digital evidence: Challenging the presumption of reliability. *Journal of Digital Forensic Practice*, 1(1), 19–26. <https://doi.org/10.1080/15567280500541421>
- Venville, N. (2015). *A review of contextual bias in forensic science and its potential legal implications*. ANZPAA-NIFS. <http://netk.net.au/Psychology/Psychology14.pdf>
- Vincze, E. A. (2016). Challenges in digital forensics. *Police Practice and Research*, 17(2), 183–194. <https://doi.org/10.1080/15614263.2015.1128163>
- Virgo, P. (2021, 25 May). Lessons from the Post Office Horizon case. *Computerweekly.com*. <https://www.computerweekly.com/blog/When-IT-Meets-Politics/Lessons-from-the-Post-Office-Horizon-Case>
- Ward, E. D. (2021). *The influence of mobile technology advancements on digital forensics investigations practices and procedures: A generic qualitative inquiry*. PhD thesis, Capella University.
- Watson, D. L., & Jones, A. (2013). *Digital forensics processing and procedures: Meeting the requirements of ISO 17020, ISO 17025, ISO 27001 and best practice requirements*. Newnes.
- Weizman, E. (2014). Introduction: Forensics. In E. Weizman (Ed.), *Forensics: The architecture of public truth* (pp. 9–32). Sternberg Press.

- Westera, N. J., Kebbell, M. R., Milne, B., & Green, T. (2016). The prospective detective: Developing the effective detective of the future. *Policing and Society*, 26(2), 197–209. <https://doi.org/10.1080/10439463.2014.942845>
- Williams, R., & Weetman, J. (2013). Enacting forensics in homicide investigations. *Policing and Society*, 23(3), 376-389. <https://doi.org/10.1080/10439463.2012.703200>
- Wilson-Kovacs, D. (2019). Effective resource management in digital forensics: An exploratory analysis of triage practices in four English constabularies. *Policing*, 43(1), 77–90. <https://doi.org/10.1108/PIJPSM-07-2019-0126>
- Wilson-Kovacs, D. (2021). Digital media investigators: challenges and opportunities in the use of digital forensics in police investigations in England and Wales. *Policing*, 44(4), 669-682. <https://doi.org/10.1108/PIJPSM-02-2021-0019>
- Wilson-Kovacs, D., Rappert, B., & Redfern, L. (2021). Dirty work? Policing online indecency in digital forensics. *The British Journal of Criminology*, 62(1), 106–123. <https://doi.org/10.1093/bjc/azab055>
- Windelband, W. (1998). History and natural science. *Theory & Psychology*, 8(1), 5–22. <https://doi.org/10.1177/0959354398081001>
- Zahadat, N. (2019). Digital forensics, a need for credentials and standards. *Journal of Digital Forensics, Security and Law*, 14(1), 1–14. <https://doi.org/10.15394/jdfsl.2019.1560>
- Årnes, A. (2018). Introduction. In A. Årnes (Ed.), *Digital forensics* (pp. 1-12). Wiley.

PART TWO

Article 1: Sunde, N., Dror, I. E. (2019). Cognitive and human factors in digital forensics: Problems, challenges, and the way forward. *Digital Investigation*, 29, 101-108.
<https://doi.org/10.1016/j.diin.2019.03.011>.



Cognitive and human factors in digital forensics: Problems, challenges, and the way forward

Nina Sunde ^{a, *}, Itiel E. Dror ^b

^a University of Oslo, Norwegian Police University College, Norway

^b University College London, UK

ARTICLE INFO

Article history:

Received 4 February 2019

Received in revised form

24 March 2019

Accepted 26 March 2019

Available online 29 March 2019

Keywords:

Digital forensics

Digital investigation

Digital evidence

Forensic science

Cognitive bias

Human error

Human factors

Expert decision making

ABSTRACT

Digital forensics is an important and growing forensic domain. Research on miscarriages of justice and misleading evidence, as well as various inquiries in the UK and the US, have highlighted human error as an issue within forensic science. This has led to increased attention to the sources of cognitive bias and potential countermeasures within many forensic disciplines. However, the area of digital forensics has yet to pay sufficient attention to this issue. The main goal of this article is to contribute to a more scientifically sound digital forensics domain by addressing the issues of cognitive bias as a source of error. In this paper we present an analysis of seven sources of cognitive and human error specifically within the digital forensics process, and discuss relevant countermeasures. We conclude that although some cognitive and bias issues are very similar across forensic domains, others are different and dependent on the specific characteristic of the domain in question, such as digital forensics. There is a need for new directions in research with regard to cognitive and human factors in digital forensics.

© 2019 Elsevier Ltd. All rights reserved.

Introduction

With the extensive and growing use of technology in everyday life (both by law enforcement and by criminals), the importance of digital forensics and the reliance on digital evidence will continue to grow. Therefore, it is important to consider how this forensic discipline can be made as robust and reliable as possible. Digital forensics began as an ad hoc, task oriented practice performed by computer professionals with no formal processes, tools or training (Garfinkel, 2010). In recent years there has been an advance within digital forensics towards a more scientifically sound handling of the evidence by enhanced focus on quality management (Page et al., 2018), error mitigation, tool testing and verification methodologies (Casey, 2011; ENFSI, 2015; Pollitt et al., 2018; SWGDE, 2018). Digital forensics is now recognized as a discipline within forensic science by a variety of organisations (e.g., the Australian National Institute of Forensic Science (NIFS), the European Network of Forensic Science Institutes (ENFSI) and the United States National Institute of Standards and Technology (NIST)).

Within the digital forensics domain the term “forensically sound” is used extensively (e.g. a search on Google Scholar on “digital evidence” + “forensically sound” returns 1610 hits). An analysis of how the term was understood in the digital forensics literature, linked the quality of which the digital forensics process was conducted to the admissibility of the digital evidence in the court of law (McKemmish, 2008). Forensically sound was defined as: “The application of a transparent digital forensics process that preserves the original meaning of the data for production in a court of law” (McKemmish, 2008, p. 10). Forensically sound reflects a threshold of minimum requirements for court presentation, inherently accepting that using only rigid and scientifically sound procedures for producing the evidence is often unachievable within the digital forensics domain. However, the quality threshold for presenting digital evidence in court should not be confused with the goal of advancing the domain towards a more *scientifically sound* domain, by adopting scientifically sound procedures for handling digital evidence which take both technological and human factors into account.

Like other forensic disciplines, and science in general, there are uncertainties, limitations, vulnerabilities, and potential for error. It is known from evaluations of wrongful conviction cases in the US, Australia, and England and Wales that from 23% to 60% of the cases

* Corresponding author.

E-mail addresses: nina.sunde@phs.no (N. Sunde), i.dror@ucl.ac.uk (I.E. Dror).

included flawed or exaggerated forensic evidence contributing to the convictions of innocent people (Dioso-Villa, 2012, 2015; Garrett, 2011; Garrett and Neufeld, 2009; National Registry of Exonerations, 2015; Smit et al., 2018).

Digital evidence is perceived as reliable and correct by many in the legal community (Van Buskirk and Liu, 2006). This should be a concern when considering the results of a recent analysis of the quality management (QM) procedures in digital forensics performed by Page et al. (2018). They compared the QM procedures between bodily fluids and DNA, fingerprint and digital forensics in the UK, and found that digital forensics operates with the least robust QM procedures.

We first distinguish between errors which arise from technology, for example errors in timestamps or loss of data (e.g. Casey, 2002; Ekfeldt, 2016; SWGDE, 2018), and those related to the lack of scientific data, research and rigor (SWGDE, 2018). In addition, and in contrast to the technological and scientific issues, there are also errors that arise from cognitive and human factors. These cross all forensic domains (as well as other expert domains). Within forensic science, there has been an increased attention towards cognitive and human factors as a source of error, with a growing body of research in this area (see below, section 2.1).

Within the digital forensics domain, there has been a movement from perceiving tools and technology as the main instruments in the digital forensics process towards greater acknowledgement of *the human* as an important instrument for examining digital evidence (e.g. ENFSI, 2015; Pollitt et al., 2018). However, as of today this issue has not been sufficiently addressed through research within this domain. When addressing human factors and human error, it is important to look into the cognitive factors that shape perception and interpretation and how they may impact the experts' decision making.

A critical element is cognitive bias, understanding where it may come from, and how to control and minimize it. Pohl (2016) explains cognitive bias as a cognitive phenomenon which reliably deviates from reality, occurs systematically and involuntarily, and is difficult or impossible to avoid by mere willpower. The main goal of this article is to contribute to a more scientifically sound digital forensics domain by addressing the issues of cognitive bias as a source of error. In this paper we examine the possible human sources of errors in the digital forensics process, and specifically the role of the human experts in making forensic decisions within this area. The research question addressed here is: When handling digital evidence through the digital forensics process; when is the digital forensics practitioner (DFP) vulnerable to cognitive bias, and what measures could be relevant and effective to mitigate bias for this specific domain?

To answer this question, we first present relevant research concerning bias and human error in forensic science. An analysis of cognitive and human error within the digital forensics process is then conducted, and some relevant countermeasures are discussed. We conclude with the need for future research and new directions with regard to cognitive and human factors in digital forensics.

Cognitive and human factors in forensic science

For decades, there was little research or attention to the human examiner in any forensic domain. This has changed, and over the past decade there has been much attention and research on bias and other cognitive factors in many forensic disciplines. However, the area of digital forensics has remained as one of the few forensic disciplines that has not sufficiently researched or given much attention to this issue. Since the digital forensics discipline shares many core forensic processes with other forensic disciplines (Pollitt et al., 2018), research from such disciplines are a good and relevant

starting point for exploring the issue of cognitive and human error.

Research on bias and forensic science

For the disciplines dealing with identification, for example DNA and fingerprint analysis, there is a solid research foundation showing that contextual information can distort the forensic analyst's judgement. Dror and Hampikian (2011) tested whether DNA experts could be biased by the details of the case. After a gang rape, evidence in form of a mixture of DNA from multiple contributors was secured as evidence and analysed. This analysis required some judgement and interpretation from the forensic analyst, as is often the case in DNA mixture analyses. One of the suspects testified against other suspects as part of a plea bargain, and this was known to the two analysts who examined the DNA evidence. Both analysts concluded that the suspect in question *could not be excluded* from being a contributor to the DNA mixture. To examine the possible biasing effect of this irrelevant contextual information (the testimony against the suspect), Dror and Hampikian (2011) presented the original DNA evidence to 17 experts, but excluded the irrelevant contextual information. The vast majority of experts did not agree with the original examiners: Only 1 of the 17 agreed with the analysts who were exposed to the biasing information. Without the irrelevant biasing context 16 examiners did not conclude that the suspect could not be excluded.

Within the fingerprint domain, Dror and Charlton (2006) presented fingerprint experts with evidence they had examined before (without their knowledge), but did so with new irrelevant contextual information. Their data demonstrated that fingerprint experts' decisions could be biased by irrelevant contextual information (see also Dror et al., 2006; Dror and Rosenthal, 2008). These findings have now been replicated by Stevenage and Bennett (2017) who found that knowledge of the results of a DNA test affected the decisions in fingerprint matching tasks, as well as by Smalarz et al. (2016) who found that even criminal stereotypes can bias forensic evidence analysis.

The biasing effect of irrelevant contextual information has also been found in research within other forensic domains, such as bloodstain pattern analysis (Taylor et al., 2016), arson investigation (Bieber, 2012), forensic pathology (Oliver, 2017), and crime scene investigation (van den Eeden et al., 2016). Similar effects have also been found in a number of studies within the domain of forensic anthropology (Nakhaeizadeh et al., 2014, 2018). For a summary of much of this research across forensic domains, see (Dror, 2016).

Compatible with a huge body of research in many scientific domains, the empirical research in various forensic disciplines has predominantly shown that human examiners do not always produce consistent results, and that they are susceptible to cognitive bias. In the context of a criminal investigation, biased conclusions may lead to errors in the fair administration of justice.

Research on misleading forensic science evidence

Errors originating from several forensic disciplines have caused or contributed to wrongful convictions in several countries (for an overview, see Smit et al., 2018). This issue has been subject to several examinations, aimed at trying to establish the causes that underlie miscarriages of justice. According to The National Registry of Exonerations, false or misleading forensic evidence were found in 23% of the 1600 individual exonerations in US from 1989 to 2015 (National Registry of Exonerations, 2015). Garrett and Neufeld (2009) looked at 137 cases where innocent persons were convicted for serious crimes such as rape or murder, and later exonerated by post-conviction DNA testing. They found that in 60% of these cases forensic evidence contributed to the wrongful

conviction. In Australia, [Dioso-Villa \(2012, 2015\)](#) looked into 71 cases of wrongful conviction, and found that 31% of the cases involved forensic errors or misleading forensic evidence. In the UK, a study by [Smit et al. \(2018\)](#) found that forensic evidence (including digital evidence) was an issue in 32% of the 235 cases they looked at in which the Court of Appeal quashed a conviction. With regards to digital evidence in particular, this evidence type is only highlighted in the UK study. Smit et al. identified 4 individual cases where presentation of CCTV recordings, SIM-card, DVD content, and web content were the misleading elements (personal communication with N. Smit, 5th of April, 2018).

To summarize, research from several forensic disciplines, as well as real criminal cases, has demonstrated that experts conducting forensic examination that involve subjectivity, interpretation or opinion are susceptible to cognitive bias. Based on the research in other forensic disciplines, as well as research in many other expert domains, and understanding of human cognitive processing, there is no reason to believe that DFPs are more reliable “instruments” than other forensic experts, nor that they are immune to bias. However, the nature of the activities and processes when analysing digital evidence differs from other disciplines, and therefore we need to examine the unique situations in digital forensics with particular risk of bias, and identify ways to specifically combat them.

In the next section we examine the various sources of bias in relation to the digital forensics process as described by [Flaglien \(2018\)](#), with the phases of identification, collection, examination, analysis and presentation. Within the digital forensic process, both investigative and forensic processes may be carried out, and the issues raised in this paper apply to both. Experiences from case-work will be used to illustrate and demonstrate the issues, including data from interviews of Norwegian DFPs and criminal detectives involved in investigations with digital evidence.

A taxonomy of sources of bias that may affect forensic decisions contextualized within the digital forensics process

The seven level taxonomy ([Fig. 1](#)) elaborates on several factors that could lead to bias, each of which entails different countermeasures ([Dror, 2017](#)). It is important to emphasize and clarify that we are talking about *cognitive bias*, and not intentional biases or errors that arise from lack of training, motivation or incompetence.

The cognitive biases arise from how the brain processes information ([Nickerson, 1998](#)). Cognitive biases are sometimes mistaken as ethical issues. This is not correct, since they are the result of computational tradeoffs occurring in the brain, and not the result of intentional or conscious acts ([Dror, 2014](#)). Experts are not immune to such biases ([Dror, 2011](#)), but they are mostly unaware of them due to the *bias blind spot* ([Pronin et al., 2002](#)), which has recently been demonstrated specifically within forensic science ([Kukucka et al., 2017](#)). Below we examine each of the levels from the taxonomy in [Fig. 1](#) in relation to activities conducted within the digital forensics process.

The term *digital evidence* covers both digital devices as well as relevant pieces of information (data) found on the devices. E.g. a collected digital device such as a laptop or a smartphone will often, by itself, be considered as evidence. During the digital forensics process, the information on the digital device is analysed, and information such as images, documents, chat conversations etc. will be documented, and constitutes separate pieces of evidence.

Cognitive architecture and the brain

The base level of the taxonomy, at the very bottom, addresses the brain, and our cognitive architecture. The human brain has limited capacity to process all the information it is presented with, and uses several strategies to cope with this. [Zapf and Dror \(2017\)](#) describes one such strategy, *chunking*, which means to bind individual pieces of information together, so they are represented as one within our mental representation. For example, when trying to memorize a phone number, instead of memorizing the number sequence 98348758, it could be divided in three chunks; 983-48-758, making the information easier to retain and recall. Another strategy is *selective attention*, meaning that we attend to a specific piece of information while ignoring other information. *Top down processing* refers to a whole set of cognitive processes that is conceptually driven by what is already in our brains (i.e., our knowledge, expectations, hope, fears, personality, prior experiences, etc.). The top-down processing uses context to make sense of the incoming information ([Lindsay and Norman, 2013](#); [Zapf and Dror, 2017](#)). These and other cognitive strategies facilitate efficient information handling to a person in everyday life, and are used more as we gain experience and expertise. These mechanisms are very helpful, useful and needed (and mostly out of our control

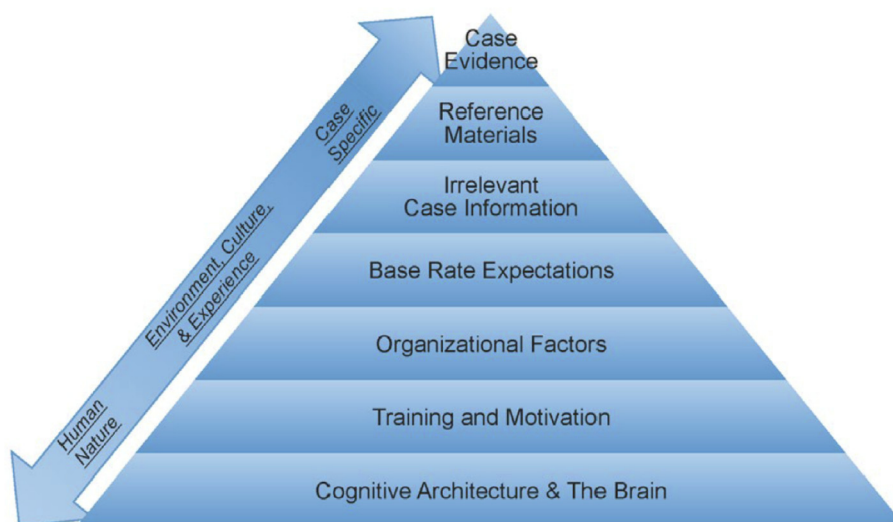


Fig. 1. Influences that might interfere with accurate observations and inferences in forensic decision making ([Dror, 2017](#)).

and awareness), but they can also degrade our impartiality and cause bias. Contrary to what many believe, in various ways an expert is actually more susceptible to bias compared to novices due to the underpinning of expertise (Dror, 2011).

The list of biases described in research literature is quite substantial. We will focus on three biases that are particularly relevant to the work of a forensic examiner. *Anchoring bias* is due to information we encounter first, making it more influential than information that is received later (Tversky and Kahneman, 1974). Once we get the first piece of information, we may fixate, escalate commitment and can have tunnel vision. *Availability bias* makes us overestimate the likelihood of an outcome based on our ability to recall similar instances (Tversky and Kahneman, 1973). In the process of making sense of what is observed, a hypothesis will often evolve (often subconsciously at first) (Kahneman, 2011), leading to *confirmation bias* (Nickerson, 1998), which is the tendency to:

- a) seek information that supports the hypothesis
- b) interpret new information in relation to the hypothesis
- c) interpret ambiguous or neutral information as supportive to the hypothesis
- d) ignore or explain away information that contradicts the hypothesis
- e) give little weight to information that does not support the hypothesis

Cognitive bias is also impacted by emotions, such as confidence, frustration, sorrow and anger, personal responsibility, and concern about future consequences (Ask, 2013). When the biases influence the DFP, they might lead to several effects, which in turn might introduce errors in the digital forensics work.

At a physical crime scene, there is a constant flow of new information from the observations at the scene, which activates top-down processing, influenced by what the DFP already knows, has experienced before or expects from the situation. The mind is constantly trying to make sense of the situation, filling in the gaps where information is missing, giving meaning to ambiguous information, and so on. What the DFP observes during a search may affect the interpretation of the latter due to anchoring and other biases. During all the phases of the digital forensics process the hypothesis under consideration has profound effects on the DFPs' observations and conclusions. If a person is suspected of committing a criminal act, the confirmation bias might lead to an aggregated focus within the investigative or forensic tasks on finding information consistent with the guilt hypothesis, while a tendency to overlook or explain away information that contradicts it, such as exonerating information or information indicating mitigating circumstances.

Training and motivation

Moving up the pyramid (Fig. 1), the next three levels relate to the environment, culture and experience. These are about our personal nature, our motivations and preferences, and are developed throughout our lives. They affect how we observe, reason and make decisions (Balcetis and Dunning, 2006). Motivation is a source of erroneous decision making in criminal investigation (Ask and Granhag, 2005). Pre-existing attitudes and with whom the individual identifies might also sway observations and conclusions (Neal, 2016).

The DFP often will come across sexual crimes against children, since the evidence in such cases often is digital images, videos or online communication. The pre-existing attitudes of the DFP may affect the analysis of digital evidence. A motivation for "saving" a child by finding evidence against a perpetrator who has harmed

children in the past might be a strong influential factor on their analysis.

Concerning the issue of training, digital forensics has roots in computer science, physics (electronics) and mathematical theory (ENFSI, 2015). Within these sciences and theories, the role of the cognitive processes and the human brain are usually not given much (if any) attention in the curriculum. The typical DFP will most often lack knowledge of cognitive mechanisms, biases and relevant countermeasures. This was apparent in the interviews of Norwegian DFPs with civilian backgrounds (Sunde, 2017). The interviews showed that they had no formal knowledge of bias and relevant countermeasures from their educational background or professional training. An important issue concerning human error and the digital forensics work is the lack of formal qualification and certification requirements. Page et al. (2018) found that compared to the DNA and fingerprint domains, the digital forensics discipline "is operating under arguably less rigorously defined standards, practitioner governance and evidence validation procedures" (Page et al., 2018, p. 2).

Organisational factors

The next level of the taxonomy (Fig. 1) is about organisational and cultural factors, which implicates social interactions, identification with ideologies and organizations, and communication with others. The words we use, the terminology, vocabulary and jargon can cause errors in how we interpret and understand information (Zapf and Dror, 2017). The information flow between different systems, organisations and entities in an organisational structure may also affect the human examiner and bias their work. Most concerning is how bias can cascade and snowball between different, supposedly independent, elements involved in criminal investigations (Dror, 2018).

The overall organization of the digital forensics work, as well as cooperation and communication procedures are relevant aspects to discuss in relation to bias. How digital forensics work is organised varies a lot, from organisations outside law enforcement providing digital forensics as a service, to digital forensics integrated within the law enforcement organisation. Norway and Finland are examples of the latter (Leppänen and Kankaanranta, 2017; Sunde, 2017). In Norway, the DFPs often work in close cooperation with both the investigation team and the prosecution (Sunde, 2017). The defence may hire private DFPs to do own analysis, and to challenge the relevance or reliability of the digital evidence presented by the prosecution in court. Research from field studies and experimental studies has showed that whether the forensic experts believed they were working for the prosecution or the defence affected their forensic conclusions (Murrie et al., 2009). This has been termed *adversarial allegiance*. This implies that although the DFP is obliged to be impartial and objective during investigation and trial, whether DFPs are working for the prosecution or the defence may affect how they observe and what they conclude about digital evidence, and how they present the evidence in court.

In relation to cooperation and communication procedures, integrating and combining digital forensics within the investigation team is sometimes regarded as an advantage. The experience from a project in the Oslo Police District was that close cooperation between the DFP and the tactical investigation team led to increased motivation, knowledge exchange and efficient communication in the investigation of sexual crimes against children (Hansen et al., 2017). In Finland, close cooperation was highlighted as a success factor behind a cost-efficient investigation of computer-integrity crimes (Leppänen and Kankaanranta, 2017). However, a downside of close cooperation between the DFP and the tactical investigation

team is the flow of biasing irrelevant information within the team, such as the results of DNA analyses, witness statements, eyewitness identifications or whether the suspect has confessed to the crime, which might lead to bias cascade and bias snowball (Dror, 2018).

Whether the forensic work is done within law enforcement or by independent laboratories, either way, the critical question is whether the forensic examiners had independence of mind during their work. Organizational factors and biasing influencing are not limited to law enforcement, they exist everywhere. Several people have tried to develop integrated models for digital forensics and criminal investigation (e.g. Ciardhuáin, 2004; Hunton, 2011; Sun et al., 2015) but none of these has addressed organisation structures (Leppänen and Kankaanranta, 2017) or taken the issue of bias mitigation into particular consideration.

Base rate expectations

The next level in the taxonomy is base rate expectations, which is about how past regularities bring about an expectation to what will be found in a new case. These are expectations for the current case based on past experience or from work with other cases in the past (Dror, 2017). Of course, the past cases can be very helpful, but they can also be misleading. This is a bias, because the impact has nothing to do with the current case at hand.

Within the identification phase of the digital forensics process, the decision about seizure may be regarded as a particularly vulnerable to bias. This phase involves search for, recognition and documentation of potential digital evidence (ASTM, 2015). Before the examination, the digital device is often a “black box” where information that could be relevant to the investigation is hidden. There are available tools such as hardware and software write blockers that enables the DFP to access and preview information from the device without altering the data (triage). However, this is not always possible due to missing tools, expertise in using the tools, or conditions on the search scene. Thus, the decision about seizure must often be made under some level of uncertainty, and base rate expectations might affect and bias the decision.

Since the base rate expectations arise from the examiners' former experience with other cases (Dror et al., 2015), having experience with not finding anything of interest on a particular type of device might result in a decision not to seize the device. There are several other important decisions to be made throughout the digital forensic process. In the collection phase the decision about how to collect (e.g. live or pull-the-plug) might also be influenced by expectations from former experience about what was successful or what effort did not pay off. In the examination phase the choice of processing software might be conducted due to the expectations about what will be found, and the base rate expectations could also affect how much effort the DFP dedicate to unpacking compressed, archived or decrypted files.

The base rate expectations are formed by previous casework, and they could create expectancies about conclusions and outcomes prior to the analysis, and may also affect what traces to look for during the analysis. This is in many ways what makes an expert, but it can also be misleading and bias them in the wrong direction when the specific case in front of them is different than previous cases. The base rate expectations degrade various aspects in the decision making process, e.g., the ability to look for, and find, critical information based on the case at hand. When the case at hand is comparable to previous cases the base rate expectations will not cause problems – they may even be helpful. However, when the case at hand has unexpected critical information, it may be overlooked and missed.

Irrelevant case information

The four previous levels are independent to the case at hand, but the three top levels of the taxonomy relate to the specific case under investigation. The case specific biasing factors are: irrelevant case information, reference materials and actual case evidence. Irrelevant case information, is about the contextual information concerning the particular case, but which is irrelevant for the referral question (i.e., not relevant to the experts work and expertise). One of the most potentially biasing considerations at this level involves the inferences made by others (Zapf and Dror, 2017). Irrelevant case information may cause *bias cascade* or *snowball effects* (Dror, 2018). The cascading effect is when a bias cascades from one stage to another within the process of handling the evidence, e.g., from the collection to the analysis stage. The snowball effect is when the bias increases as irrelevant information from a variety of sources (e.g., erroneous identification of suspect by witness, arrest of suspect, confession, etc.) is integrated and influences each other. As more people are influenced, they in turn influence others, and hence the bias increases, and gathers more momentum as more people are affected by it and then affect others, hence the snowball effect.

Within digital forensics the issue of irrelevant case information has several aspects, regarding where it occurs and how it could be managed. Irrelevant case information regarding digital evidence may be the formal case file, but also includes informal knowledge concerning the case from those involved in the investigation. This could be factual information, e.g. whether the suspect has confessed or results from analysis of other lines of evidence. Irrelevant case information could also be subjective information such as sharing of personal opinions about which offence hypothesis they consider the most likely, how they feel about the suspect, or their sympathy for the victim. If the case gets attention from the media, further biasing irrelevant case information may be found there. Due to insufficient quality management systems (Page et al., 2018), information irrelevant to the work of the DFP may be poorly managed. This is particularly problematic when the DFP is integrated within the criminal investigation team. When the DFP is conducting the collection on the search scene and then does the analysis, the bias cascades, and irrelevant contextual information from the case and other lines of evidence may influence the analysis.

In contrast to fingerprint and firearm analysis, handwriting examination, and many other forensic domains, digital evidence analysis often *requires* more context. For example in the fingerprint domain, the expert may conduct the identification process with limited contextual information, restricted to only how the prints were lifted and developed, from what surface the prints were collected, etc. In the digital forensics domain the situation is quite different. In the past digital storage devices had limited capacity, and it was possible to examine them from start to end. Now, due to the amount and complexity of data, this approach is often unfeasible. As reflected in an interview of a DFP, such situations require the use of different search strategies, and hence it was important to also report not only the result, but also what was searched for (Sunde, 2017). This is done to avoid misunderstandings about negative findings. For example, a search for images of sexual abuse of children with negative result, a null finding, could not exclude the possibility that such images were on the digital device but were just not found. Such strategies often require some context to be sufficiently customised for the particular information requirement, and a completely blinded examination of a hard drive would be inefficient and lengthy.

The case file might contain relevant information for a targeted analysis of the digital seizure. However, there would also be

irrelevant information in the case file. The case information available to the DFP would affect what they would look for, and how they would perceive the relevance of the potential evidence they observe. For instance multiple Norwegian DFPs said that they would normally have access to the case file, and would sometimes read the suspect interviews prior to their analysis (Sunde, 2017). Depending on the task the DFP is required to do, the information from this interview could be relevant or irrelevant, as well as biasing.

Reference material

Several of the forensic disciplines conduct tasks involving reference material. In an investigation involving firearms, the forensic analyst might be tasked to compare the breech face mark found on a cartridge at a crime scene with a test fire cartridge from a weapon found at the house of the suspect. The reference material itself could be biasing, especially when the forensic examiners work backwards, from the suspect to the evidence. In digital forensics, reference material could be used when, for example, the DFP is tasked to determine whether a particular file, such as an image, a document or a video is present on a computer. The task would be performed by calculating a checksum and searching for a match. In such occasions there would be very little need for interpretation by the DFP, and hence a low risk of a biased decision.

However, there are tasks where the reference materials could bias the interpretation of the actual evidence. The digital evidence is often used for cross-checking information retrieved through the investigation, for example from police interviews. When reading the interview, the DFP would often be exposed to more (biasing) information than what is necessary to conduct the cross-checking task. To minimize potential bias the DFP does not need the full statement, nor all the details, but some key information from the statement. The reference material constitutes a *target driven bias* where it, rather than the actual evidence, is guiding the cognitive process. For example, in DNA analyses, this occurs when the suspect's profile influences how the biological material from the crime scene – the actual evidence - is interpreted (Jeanguenat et al., 2017). This could occur in digital forensics, for example, when the DFP has an image of a child who has been kidnapped and now is suspected to be present in video. This may influence how the image of a child in the digital evidence is interpreted to fit the target image.

Case evidence

On the very top of the taxonomy is the case evidence itself. This category includes different types of evidence that was obtained through the investigation of the crime. Whether the piece of evidence could be a source of bias depends on whether it inherently includes biasing information within the evidence itself. A latent fingerprint or a firearm cartridge by itself does not convey any biasing information (e.g., a cartridge, by itself, does not say whether this was a misfire from a firearm, or if it was involved in a homicide). In contrast, a seized digital device might by itself be biasing, since it may have information such as nametags, stickers or other characteristics which could affect the observations and decisions of the DFP. Furthermore, the *evidence in itself*, in the form of information found through the examination of the data on the digital device, could be biasing. Examples of such could be images of sexual abuse, chat logs about the planning of a terrorist attack, or an internet search log conveying search phrases about how to get away with murder or how to spike a drink with drugs to commit a rape.

To summarize, the analysis of the seven level taxonomy in

relation to digital forensics indicates that human cognition is involved to a great extent in the work of the DFP. Digital forensics has specific sources and ways that can cause bias. Conducting the digital forensics process encompasses many observations, judgments and conclusions which involve subjectivity, interpretation and opinion by the DFP. Hence, none of the phases of the digital forensics process should be considered safe against cognitive bias or other errors caused by human cognition. In the following section we will look into ways in which bias may be mitigated, based on research and examples from other forensic disciplines.

Bias mitigating countermeasures with relevance to digital forensics

Every examination and evaluation involving a human judgement is prone to bias (Zapf and Dror, 2017). However, research has identified several promising cognitive practices and operating procedures for mitigation of biasing effects, and their relevance and feasibility for the digital forensics domain should also be considered. First and foremost, to understand how observations and conclusions may be biased, we believe that relevant cognitive psychology should be included in the *training* of the DFP. To counter the bias blind spot, the training should be practical and scenario-based, allowing the DFPs themselves to understand and experience how bias could occur without them noticing it in the normal work situation within the digital forensics domain.

The DFP, similar to criminal investigations and scientific inquiry in general, should conduct their examination by testing and eliminating *multiple*, and preferably *competing* hypotheses. A similar bias mitigation strategy is used in well managed line-ups. Instead of presenting the suspect alone, the eye-witness is exposed to multiple possible offenders together with the suspect (Stebly et al., 2003; Wells et al., 1998). The hypotheses should be defined on either offence, activity or source level, (Cook et al., 1998), and constructed on the basis of relevant information from the case. A systematic collection and consideration of relevant data to test alternative hypotheses and a careful documentation of which piece of information is being evaluated, and whether it weakens or strengthens the alternative hypotheses may help to minimize confirmation bias. This is supported by the results from experiments conducted by Rassin (2018), who found that using a pen-and-paper tool for evaluation of evidence against two competing hypotheses could reduce tunnel vision.

Irrelevant case information as a source of bias may be eliminated by thorough *context management* that can control what information is given, when, to whom, and how (Dror, 2014; Dror et al., 2015; Stoel et al., 2014). This issue, if understood, can be managed through the use of case managers (Dror, 2014), who should know everything about the case, and interact with the investigation team throughout, but they would not carry out the actual forensic analysis.

Relevant case information could be biasing, and a measure to simply avoid or isolate it is not necessarily the right course of action due to the role it may play in examination and evaluation of the evidence. Hence, a detailed analysis needs to be done to determine how to handle such situations (Dror, 2012). Rather than just blinding the examiner to information, there are approaches that control when the information is given. Dror et al. (2015) developed the Linear Sequential Unmasking (LSU) procedure, aimed at mitigating bias caused by relevant case information through *exposure control*. The LSU procedure is aimed to ensure the cognitive process goes in the right direction, from the evidence to the suspect, by regulating when the examiner should be exposed to the biasing case relevant information, and provides restrictions to when and to what degree they may go backwards and change their analysis

decisions. An important principle of LSU is that information is given *when* it is needed, hence it is not about blinding to information, but rather at optimising the sequencing.

In addition to blinding and controlling when information is given, another approach is to *compartmentalise* the work, so as to avoid bias cascade and cross cognitive contamination (Dror, 2014). In digital forensics, this could be done by separating the task, so that the analysis is assigned to another DFP than the one collecting the evidence at the crime scene. This way, the DFP that was exposed to a lot of irrelevant contextual information from the crime scene is not the DFP who conducts the subsequent analysis back at the forensic laboratory.

As bias can never be totally eliminated, procedures to *uncover* cognitive or human error will also be necessary, so as to be transparent. There are few formalized quality assurance procedures within digital forensics, such as verification or peer review, but these have to be done correctly from a cognitive point of view, e.g., that verification is done blindly. Peer review is mentioned in some guidelines and standards (e.g. ENFSI, 2015; SWGDE, 2018), but the detail in how the measure should be carried out varies from just mentioning the measure with no elaboration (e.g. SWGDE, 2018), to a more detailed description on how peer review should be conducted and documented (ENFSI, 2015). The efficiency of the peer-review procedure depends partly on the ability to reduce bias with the assessor conducting the peer review.

Peer review, or case review, is typically different from verification. Verification may be done by totally repeating the analysis by a second independent examiner to see whether they arrive at the same conclusions (Ballantyne et al., 2017). Verification hinges and depends on repeatability, which is an issue across forensic domains, including fingerprinting (Dror et al., 2011; Ulery et al., 2012) and DNA (Barrio et al., 2018; Butler et al., 2018; Dror and Hampikian, 2011). According to Page et al. (2018), proper verification does not seem to be frequently used within the digital forensic discipline. To help combat bias, the verification of digital forensics decisions should be a controlled process in which blind and double-blind procedures are used whenever possible. Such procedures would require that the verifier is not informed of the initial conclusion; if possible, that the verifier does not know who the examiner was; and that the examiner does not select the verifier (Kassin et al., 2013). To maximize independence in the verification process it is also advised, if possible and warranted, to do cross-laboratory verifications of the forensic work (Koppl et al., 2008). To avoid the base rate problem, verification should not only be undertaken on positive results, but also include verification of negative results (Dror, 2014).

Bias mitigating measures are sometimes applicable, but other times they are neither practical, possible nor advisable. Regardless of the measures that are taken to minimize bias, it is important that the DFP is *transparent* about their work. They need to provide detail in the documentation and presentation in court about what context they knew, who gave it to them, when, and why. This is particularly important regarding irrelevant contextual information they were exposed to. Hence, if bias cannot be minimized, then at least there needs to be transparency to its possible presence, which in turn will make it possible for others to assess the reliability of the outcome, and to develop new ways to minimize bias.

Conclusion

Within forensic science, there has been increased attention on cognitive and human factor as a source of error, and quite an extensive research foundation from many forensic disciplines, which have shown that the forensic expert is susceptible to bias when making decisions. In this paper we have addressed the

research question: When handling digital evidence through the digital forensics process; when is the DFP vulnerable to cognitive bias, and what measures could be relevant and effective to mitigate bias for this specific domain?

There is a lack of research on cognitive bias within the digital forensics domain. The analysis of the digital forensics work within the seven level taxonomy of the sources of bias showed that the cognitive processes and judgement of the DFP are susceptible to many sources of bias during the digital forensics process. Given that the digital forensics domain is operating under less rigorously defined standards, practitioner governance and validation procedures compared to other forensic disciplines (Page et al., 2018), there should be a concern in relation to the risk of human error. Research on miscarriages of justice and misleading evidence has shown that human error within forensic science is not just a theoretical issue, but a real and serious problem in the justice system.

There are several research based approaches aimed at mitigating cognitive and human error in forensic examinations such as a strict scientific approach, context management, exposure control and compartmentalization of work, and some variations of these will be applicable to digital forensics. However, there are factors that are unique to the nature of digital evidence and the work of the DFP which underpins the need for research on which situations or activities are most vulnerable in relation to cognitive bias, and which are the most effective countermeasures, specific to this domain. The Hierarchy of Expert Performance framework (Dror, 2016) could be used for unpacking and measuring forensic expert performance in relation to susceptibility to bias and reliability issues through empirical research. Further research on cognitive and human error within digital forensics would contribute to a more scientifically sound discipline.

Declaration of interests

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.

Acknowledgement

We want to thank Helene O. I. Gundhus, Johanne Y. Dahl, Fergus Toolan, and the anonymous reviewers for their valuable comments on an earlier draft of this paper. This paper was funded by a grant from the Norwegian Police University College awarded to the first author for pursuing a PhD.

References

- Ask, K., 2013. Bias: Fejl og faldgruber i efterforskning. In: Hald, C., Vrist Rønn, K. (Eds.), *Om at opdage: Metodiske refleksjoner over Politiets Undersøgelingspraksis*. Fredriksberg: Samfundslitteratur.
- Ask, K., Granhag, P.A., 2005. Motivational sources of confirmation bias in criminal investigations: the need for cognitive closure. *J. Investigative Psychol. Offender Profiling* 2 (1), 43–63.
- ASTM, 2015. Standard Terminology for Digital and Multimedia Evidence Examination. ASTM E2916 – 13.
- Balctetis, E., Dunning, D., 2006. See what you want to see: motivational influences on visual perception. *J. Personal. Soc. Psychol.* 91 (4), 612–625.
- Ballantyne, K.N., Edmond, G., Found, B., 2017. Peer review in forensic science. *Forensic Sci. Int.* 277, 66–76.
- Barrio, P., Crespillo, M., Luque, J., Aler, M., Baeza-Richer, C., et al., 2018. GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06). Reporting conclusions: results and evaluation. *Forensic Sci. Int.: Genetics* 35, 156–163.
- Bieber, P., 2012. Measuring the impact of cognitive bias in fire investigation. In: Paper Presented at the Proceedings of the International Symposium on Fire Investigation. Science and Technology.
- Butler, J.M., Kline, M.C., Coble, M.D., 2018. NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned. *Forensic Sci. Int.: Genetics* 37, 81–94.
- Casey, E., 2002. Error, uncertainty, and loss in digital evidence. *International Journal*

- of Digital Evidence 1 (2), 1–45.
- Casey, E., 2011. *Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet*. Elsevier, Amsterdam.
- Ciardhuáin, S.O., 2004. An extended model of cybercrime investigations. *International Journal of Digital Evidence* 3 (1), 1–22.
- Cook, R., Evett, I.W., Jackson, G., Jones, P., Lambert, J., 1998. A hierarchy of propositions: deciding which level to address in casework. *Sci. Justice* 38 (4), 231–239.
- Dioso-Villa, R., 2012. Without legal obligation: compensating the wrongfully convicted in Australia. *Albany Law Rev.* 75 (3), 1329–1372.
- Dioso-Villa, R., 2015. A repository of wrongful convictions in Australia: first steps toward estimating prevalence and causal contributing factors. *Flinders Law Journal* 17, 163–202.
- Dror, I.E., 2011. The paradox of human expertise: why experts get it wrong. In: Kapur, N. (Ed.), *The Paradoxical Brain*. Cambridge University Press, Cambridge, pp. 177–188.
- Dror, I.E., 2012. Letter to the Editor—combating bias: the next step in fighting cognitive and psychological contamination. *J. Forensic Sci.* 57 (1), 276–277.
- Dror, I.E., 2014. Practical solutions to cognitive and human factor challenges in forensic science. *Forensic Sci. Policy Manag.: Int. J.* 4 (3–4), 105–113.
- Dror, I.E., 2016. A hierarchy of expert performance. *Journal of Applied Research in Memory and Cognition* 5 (2), 121–127.
- Dror, I.E., 2017. Human expert performance in forensic decision making: seven different sources of bias. *Aust. J. Forensic Sci.* 49 (5), 541–547.
- Dror, I.E., 2018. Biases in forensic experts. *Science* 360 (6386), 243. <https://doi.org/10.1126/science.aat8443>.
- Dror, I.E., Champod, C., Langenburg, G., Charlton, D., Hunt, H., Rosenthal, R., 2011. Cognitive issues in fingerprint analysis: inter-and intra-expert consistency and the effect of a 'target' comparison. *Forensic Sci. Int.* 208 (1–3), 10–17.
- Dror, I.E., Charlton, D., 2006. Why experts make errors. *J. Forensic Identif.* 56 (4), 600–616.
- Dror, I.E., Charlton, D., Péron, A.E., 2006. Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Sci. Int.* 156 (1), 74–78.
- Dror, I.E., Hampikian, G., 2011. Subjectivity and bias in forensic DNA mixture interpretation. *Sci. Justice* 51 (4), 204–208.
- Dror, I.E., Rosenthal, R., 2008. Meta-analytically quantifying the reliability and biasability of forensic experts. *J. Forensic Sci.* 53 (4), 900–903.
- Dror, I.E., Thompson, W.C., Meissner, C.A., Kornfield, I., Krane, D., Saks, M., Risinger, M., 2015. Letter to the editor - context management toolbox: a linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *J. Forensic Sci.* 60 (4).
- Ekkfeldt, J., 2016. *Om Informationsstektiskt Bevis*. Doctoral Dissertation. Juridiska Institutionen, Stockholms Universitet, Ekkfeldt, J., Stockholm.
- ENFSI, 2015. *Best Practice Manual for the Forensic Examination of Digital Technology*. ENFSI-BPM-FOT-01. Version 01 (November 2015).
- Flaglien, A.O., 2018. *The digital forensics process*. In: Arnes, A. (Ed.), *Digital Forensics*. Wiley, Hoboken.
- Garfinkel, S.L., 2010. Digital forensics research: the next 10 years. *Digit. Invest.* 7, 64–73.
- Garrett, B.L., 2011. *Convicting the Innocent*. Harvard University Press.
- Garrett, B.L., Neufeld, P.J., 2009. Invalid Forensic Science Testimony and Wrongful Convictions. *Virginia Law Review*, pp. 1–97.
- Hansen, H.A., Andersen, S., Axelsson, S., Hopland, S., 2017. Case study: a new method for investigating crimes against children. In: Paper Presented at the Proceedings of the Conference on Digital Forensics, Security and Law, Maidens.
- Hunton, P., 2011. A rigorous approach to formalising the technical investigation stages of cybercrime and criminality within a UK law enforcement environment. *Digit. Invest.* 7 (3–4), 105–113.
- Jeanguenat, A.M., Bruce Budowle, B., Dror, I.E., 2017. Strengthening forensic DNA decision making through a better understanding of the influence of cognitive bias. *Sci. Justice* 57 (6), 415–420.
- Kahneman, D., 2011. *Thinking, Fast and Slow*. Macmillan.
- Kassin, S.M., Dror, I.E., Kukucka, J., 2013. The forensic confirmation bias: problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition* 2 (1), 42–52.
- Koppl, R.G., Kurzban, R., Kobilinsky, L., 2008. Epistemics for forensics. *Episteme* 5 (2), 141–159.
- Kukucka, J., Kassin, S.M., Zapf, P.A., Dror, I.E., 2017. Cognitive bias and blindness: a global survey of forensic science examiners. *Journal of Applied Research in Memory and Cognition* 6 (4), 452–459.
- Leppänen, A., Kankaanranta, T., 2017. Cybercrime investigation in Finland. *J. Scand. Stud. Criminol. Crime Prev.* 18 (2), 157–175.
- Lindsay, P.H., Norman, D.A., 2013. *Human Information Processing: an Introduction to Psychology*. Academic press.
- McKemmish, R., 2008. When is digital evidence forensically sound? In: *IFIP International Conference on Digital Forensics*. Springer, Boston, MA, pp. 3–15.
- Murrie, D.C., Boccaccini, M.T., Turner, D.B., Meeks, M., Woods, C., Tussey, C., 2009. Rater (dis) agreement on risk assessment measures in sexually violent predator proceedings: evidence of adversarial allegiance in forensic evaluation? *Psychol. Publ. Pol. Law* 15 (1), 19–53.
- Nakhaeizadeh, S., Dror, I.E., Morgan, R.M., 2014. Cognitive bias in forensic anthropology: visual assessment of skeletal remains is susceptible to confirmation bias. *Sci. Justice* 54 (3), 208–214.
- Nakhaeizadeh, S., Morgan, R.M., Rando, C., Dror, I.E., 2018. Cascading bias of initial exposure to information at the crime scene to the subsequent evaluation of skeletal remains. *J. Forensic Sci.* 63 (2), 403–411.
- National Registry of Exonerations, 2015. *The First 1600 Exonerations downloaded*. https://www.law.umich.edu/special/exoneration/Documents/1600_Exonerations.pdf. (Accessed 20 March 2019).
- Neal, T.M., 2016. Are forensic experts already biased before adversarial legal parties hire them? *PLoS One* 11 (4) e0154434. <https://doi.org/10.1371/journal.pone.0154434>.
- Nickerson, R.S., 1998. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* 2 (2), 175–220.
- Oliver, W.R., 2017. Effect of history and context on forensic pathologist interpretation of photographs of patterned injury of the skin. *J. Forensic Sci.* 62 (6), 1500–1505.
- Page, H., Horsman, G., Sarna, A., Foster, J., 2018. A review of quality procedures in the UK forensic sciences: what can the field of digital forensics learn? *Sci. Justice* 59 (1), 83–92.
- Pohl, R.F., 2016. *Cognitive Illusions: Intriguing Phenomena in Judgement, Thinking and Memory*. Psychology Press.
- Pollitt, M., Casey, E., Jaquet-Chiffelle, D., Gladyshev, P., 2018. A Framework for Harmonizing Forensic Science Practices and Digital/Multimedia Evidence. The Organization of Scientific Area Committees for Forensic Science (OSAC), USA.
- Pronin, E., Lin, D.Y., Ross, L., 2002. The bias blind spot: perceptions of bias in self versus others. *Pers. Soc. Psychol. Bull.* 28 (3), 369–381.
- Rassin, E., 2018. Reducing tunnel vision with a pen-and-paper tool for the weighting of criminal evidence. *J. Investigative Psychol. Offender Profiling* 15 (2), 227–233.
- Smalarz, L., Madon, S., Yang, Y., Guyll, M., Buck, S., 2016. The perfect match: do criminal stereotypes bias forensic evidence analysis? *Law Hum. Behav.* 40 (4), 420–429. <https://doi.org/10.1037/lhb0000190>.
- Smit, N.M., Morgan, R.M., Lagnado, D.A., 2018. A systematic analysis of misleading evidence in unsafe rulings in England and Wales. *Sci. Justice* 58 (2), 128–137.
- Stebly, N., Dysart, J., Fulero, S., Lindsay, R., 2003. Eyewitness accuracy rates in police showup and lineup presentations: a meta-analytic comparison. *Law Hum. Behav.* 27 (5), 523–540.
- Stevenage, S.V., Bennett, A., 2017. A biased opinion: demonstration of cognitive bias on a fingerprint matching task through knowledge of DNA test results. *Forensic Sci. Int.* 276, 93–106.
- Stoel, R.D., Berger, C., Kerkhoff, W., Mattijssen, E., Dror, I.E., Hickman, M., Strom, K., 2014. Minimizing contextual bias in forensic casework. *Forensic Science and the Administration of Justice: Critical Issues and Directions* 67, 67–86.
- Sun, J.-R., Shih, M.-L., Hwang, M.-S., 2015. A survey of digital evidences forensic and cybercrime investigation procedure. *Int. J. Netw. Secur.* 17 (5), 497–509.
- Sunde, N., 2017. *Non-technical Sources of Errors when Handling Digital Evidence within a Criminal Investigation*. Master's Thesis. Faculty of Technology and Electrical Engineering, Norwegian University of Science and Technology, Sunde, N. Gjøvik.
- SWGDE, 2018. *Establishing Confidence in Digital Forensic Results by Error Mitigating Analysis*. Version: 2.0 (20 Nov, 2018).
- Taylor, M.C., Laber, T.L., Kish, P.E., Owens, G., Osborne, N.K., 2016. The reliability of pattern classification in bloodstain pattern analysis, Part 1: bloodstain patterns on rigid non-absorbent surfaces. *J. Forensic Sci.* 61 (4), 922–927.
- Tversky, A., Kahneman, D., 1973. Availability: a heuristic for judging frequency and probability. *Cogn. Psychol.* 5 (2), 207–232.
- Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. *Science* 185 (4157), 1124–1131.
- Ulery, B.T., Hicklin, R.A., Buscaglia, J., Roberts, M.A., 2012. Repeatability and reproducibility of decisions by latent fingerprint examiners. *PLoS One* 7 (3), 1–12.
- Van Buskirk, E., Liu, V.T., 2006. Digital evidence: challenging the presumption of reliability. *J. Digit. Forensic Pract.* 1 (1), 19–26.
- van den Eeden, C.A., de Poot, C.J., Van Koppen, P.J., 2016. Forensic expectations: investigating a crime scene with prior information. *Sci. Justice* 56 (6), 475–481.
- Wells, G.L., Small, M., Penrod, S., Malpass, R.S., Fulero, S.M., Brimacombe, C., 1998. Eyewitness identification procedures: recommendations for lineups and photospreads. *Law Hum. Behav.* 22 (6), 603–647. <https://doi.org/10.1023/A:1025750605807>.
- Zapf, P.A., Dror, I.E., 2017. Understanding and mitigating bias in forensic evaluation: lessons from forensic science. *Int. J. Forensic Ment. Health* 16 (3), 227–238.

Article 2: Sunde, N., Dror, I. E. (2021). A Hierarchy of Expert Performance (HEP) applied to digital forensics: Reliability and biasability in digital forensics decision making. *Forensic Science International: Digital Investigation*, 37, 301175.
<https://doi.org/10.1016/j.fsidi.2021.301175>.



Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

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

A hierarchy of expert performance (HEP) applied to digital forensics: Reliability and biasability in digital forensics decision making

Nina Sunde ^{a, *}, Itiel E. Dror ^b^a University of Oslo, Norwegian Police University College, Norway^b University College London, UK

ARTICLE INFO

Article history:

Received 7 November 2020

Received in revised form

29 April 2021

Accepted 3 May 2021

Available online 24 May 2021

Keywords:

Bias

Reliability

Digital forensics

Forensic science decision making

Human factors

Contextual information

Expert judgment

ABSTRACT

In order to examine the biasability (impact of contextual information) and reliability (consistency) of digital forensic observations, interpretations, and conclusions, 53 digital forensics (DF) examiners analysed the same evidence file. For biasability, some DF examiners were provided with contextual information suggesting guilt or innocence, while a control group received no contextual information. As per biasability, the results showed that the DF examiners' observations were affected by the biasing contextual information. As per reliability, the results showed low reliability between DF examiners in observations, interpretations, and conclusions. For improving DF work, as well as for transparency, it is important to study and assess the biasability and reliability of their decision making.

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

There is growing use of digital evidence in criminal investigations, a trend that is likely to increase with greater use of technological devices by the public and with new digital forensics (DF) capabilities. Although digital evidence is often perceived as objective and credible, there is a range of possible errors and uncertainties inherent in the evidence itself (Casey, 2002), as well as errors that can derive from the human factors involved in the processes in which the evidence is produced (Pollitt et al., 2018; Sunde and Dror, 2019). DF work involves many judgements and decisions that require interpretation and subjectivity (Sunde and Dror, 2019). This means that the quality of the outcome of the DF process – digital evidence – is dependent on cognitive and human factors, which can lead to bias and error.

DF is a relatively new and fast-changing domain in forensic science. The ever-changing landscape of technology, both of the evidence itself as well as the forensic tools to examine it, require constant adaptation and updating of methodologies and tools. This constant flux may have contributed to what may be described as a

'quality challenge' in DF. Page et al. (2019) referred to the quality regime of DF as a "wild west", where key components for quality management rarely seemed to be in place.

The DF domain faces challenges with rapid technology changes, increasing complexity, issues of accessibility, and increasing amounts of data. Research has attempted to shed light on how to produce high-quality digital evidence within this challenging context (e.g. Watkins et al., 2009; van Baar et al., 2014; van Beek et al., 2015; van Beek et al., 2020). However, the role of the human DF examiner has only been subject to research in a few studies (e.g. James and Gladyshev, 2013; Wilson-Kovacs, 2019). Specifically, performance and decisions making in terms of reliability (i.e., consistency, repeatability, reproducibility) and biasability have not been given sufficient attention. Understanding the human role, and the sources of bias and error (Dror, 2020) is crucial for developing effective quality measures and for transparency.

The research on expert performance and decision making in other forensic science domains has shown that even DNA analysts, fingerprint examiners, and forensic toxicology analysts can produce inconsistent results and are prone to bias (e.g. Dror and Hampikian, 2011; Dror et al., 2006).

To gain knowledge about potential bias and error in the DF process, it is therefore important to study DF decision making. This

* Corresponding author.

E-mail address: Nina.Sunde@phs.no (N. Sunde).

knowledge is an essential foundation for transparency and designing effective measures that can minimize error, or detect them before they cascade further into the investigation process. Invalid digital evidence is a threat to the fair administration of justice and can cause wrongful convictions (Garrett and Neufeld, 2009; Garrett, 2021) or the release of guilty culprits.

1.1. Hierarchy of expert performance

The Hierarchy of Expert Performance (HEP) (Dror, 2016 – see Fig. 1) is a useful framework for exploring and quantifying expert performance in DF. This framework has been applied to examine forensic anthropology (Hartley and Winburn, 2021), forensic interviewing (Huang and Bull, 2020), and forensic psychology (Dror and Murrice, 2018). HEP covers three perspectives of decision making: first, reliability vs biasability; second, observations vs conclusions; and third, differences between (among and across) experts vs within experts (same expert, same evidence, at different times) (see Fig. 1).

Reliability and biasability are two fundamental properties of decision making. Reliability refers to the consistency, reproducibility, or repeatability of decisions, i.e., would the same observations and conclusions be made on the same evidence. Biasability refers to the effects of task-irrelevant contextual information and other biases (see eight different sources of bias, Dror, 2020) that can impact observations and conclusions. For example, knowing that the suspect was arrested, that they confessed or even their race may influence the experts' judgements and decisions. Biasability can be explored by assessing whether such contextual information impacts decision making. Reliability in decision making can be explored by assessing whether decisions are consistent when the decision-makers analyse the same evidence with the same information basis.

Reliability should not be confused with validity. The DF examiners may produce consistent results, but this does not necessarily mean that they are valid, since a reliable tool or a repeatable process may perfectly well produce consistent, but invalid, results. While reliability is concerned with consistency, validity is about the results being correct (Christensen et al., 2014). These concepts are connected, since if there is no reliability, there are no consistent results, and thus one cannot even consider validity (if the ground

truth is unknown). Only when results are consistent, then one can examine whether they are valid, thus, reliability is a prerequisite for considering validity.

Reliability and biasability can be examined as per the conclusions reached (see levels 5–8 of Fig. 1), that is, whether they are reliable and are impacted by bias. However, expert decisions are underpinned by the observations, the perception of what the data are. HEP takes apart these different decision making components, which are often mixed (see levels 1–4 of Fig. 1). For example, differences in decisions may be incorrectly attributed to different ways of interpreting and concluding, when the different decisions actually emerge from inconsistencies in the observations of data. All of these comparisons and perspectives can be examined between examiners, or within the same examiner at different times.

While some of DF work is about presenting observed traces (observation) or conducting and presenting conclusions with formal evaluations of evidence in the light of propositions (conclusions), the result of DF work often involves presenting interpretations of the observed evidence/traces (what they mean) (Tart et al., 2019). These interpretations are inferences based on observations, measurements, or experimental testing where the DF examiner starts “connecting the dots”, and prepares the ground for a conclusion. Concerning HEP, such interpretations of observed evidence are somewhere between observations and conclusions, which is between HEP levels 4 and 5 in Fig. 1. All three levels; observation, interpretation of observation, and conclusion were explored in the current study.

1.2. Contextual information and decision making

Bias may originate from many sources, and one of these is contextual information. For biasability, contextual information can be divided into two categories: *task-relevant* and *task-irrelevant*. To minimize contextual bias, task-irrelevant information should be excluded from the forensic decision making process. Task-relevant information is necessary for forensic decision making, but should - if possible - be managed and introduced at the right time to prevent it from biasing the decision making processes (see, for example, the Linear Sequential Unmasking (LSU) procedure, Dror, 2020). Due to the biasing nature of contextual information, the forensic examiner should be transparent about what they knew and at which point in time while conducting their forensic work.

Classifying what is task-relevant and task-irrelevant can be difficult, and could also vary from case to case. However, some information would be clearly task-irrelevant, such as whether the suspect is arrested, whether the suspect has confessed, and whether the police detective thinks they are guilty or innocent.

We conducted a background investigation to explore how DF work was commissioned to ascertain whether DF examiner would normally be exposed to contextual information. Research into forensic fingerprinting has shown that examiners are exposed to task-irrelevant contextual information through the submission forms, such as whether the suspect has previous criminal convictions (Gardner et al., 2019). We extend this study to the DF domain and collected 30 submission forms from different DF units/organisations from Europe and the United States.

The forms we obtained included a variety of information, such as names of suspects, victims, gender, addresses, information about the seized device, and the place/situation from which it was collected. Also, 22 (73%) of the forms had a section where the commissioning party was requested to provide ‘information about the case’ (or comparable phrasings). None of the forms explicitly stated not to provide task-irrelevant information or provided instruction about what information should be included (or not included) in this section. The ‘information about the case’ was often

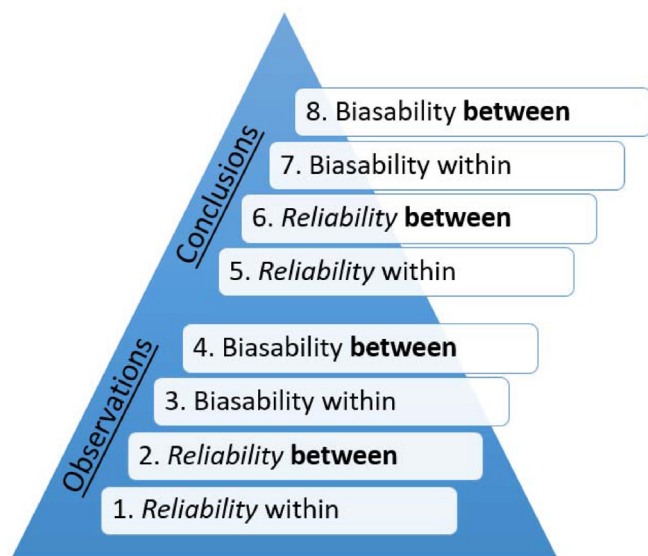


Fig. 1. Hierarchy of expert performance (Dror, 2016).

a large portion of the form, allocating much writing space (and sometimes noting to add addition if the space was not sufficient).

Eleven (over a third) of the DF units informed us that instead of, or in addition to, using a form they would also directly communicate with the commissioning party to clarify the task and 'expectations'. Thus, the submission form seemed to function as a starting point for further dialogue concerning the task and introduced no restrictions or limitations about the information exchange. The pre-study suggest that it is common to provide contextual information when DF work is commissioned and that some of this information may be task-irrelevant.

The aim of the study reported below was twofold. First, to explore whether contextual information biases the DF decision making, and second, whether the DF examiners are consistent in their decision making. Specifically, the research questions we address in this study are:

1. Are DF examiners biased by contextual information when making observations, interpretations of observations, or in their conclusions during the analysis of digital traces?
2. Are DF examiners consistent with one another when making observations, interpretations of observations, or conclusions during the analysis of digital traces?

Concerning HEP, the current study specifically examined the reliability and biasability between examiners at observation and conclusion, levels 2, 4, 6, and 8 in HEP (see Fig. 1). To address the first research question concerned with biasability in DF decision making, we gave the same evidence file to the examiners and manipulated if and what contextual information was given. A control group received no contextual information, and the other groups received contextual information suggesting either strong guilt, weak guilt/ambiguous, or innocence. The strong guilt context was clearly task-irrelevant (the suspect has confessed to the crime), while the other contexts contained both task-relevant and irrelevant information (see Appendix 1 for details). We first examined whether the contextual information biased the DF examiner's observations (HEP 4). Second, we explored whether interpretations of the observations were biased by the contextual information. And third, we examined whether their conclusions were impacted by the contextual information (HEP level 8).

To address the second part of the study concerning reliability in DF decision making, we first examined how consistent DF examiners who examined the same evidence file were (within the same contextual information), in their observations of information (HEP level 2). Second, we explored how consistent they were in their interpretations of the observed information, and third, how consistent they were in their conclusions (HEP level 6).

2. Method

2.1. Participants

An invitation letter was sent out to DF examiners through the European Union Cybercrime Task Force (EUCTF) network, Interpol digital forensics expert network, as well as DF departments within the Norwegian Police. The invitation specified that the participants should be qualified and experienced with DF work and that analysis of digital evidence in criminal investigations should be their main professional task. The participants were informed that the research aimed at gaining knowledge about factors that may increase the quality of DF investigations and the reliability of digital evidence, and that the purpose of the experiment was to explore to what degree that participants arrived at similar results, and how the results were documented.

This provided us with 65 consenting participants. Nine did not complete the experiment, and 3 were excluded from the sample (2 due to the lack of sufficient expertise in performing analysis of evidence files, and 1 participant due to not being a practitioner). The data from the remaining 53 participants (82% of the initial participants) were included in the analysis.

The DF examiners came from 8 different countries: Norway (44), India (2), UK (2), Denmark (1), Finland (1) The Netherlands (1), Kenya (1), Canada (1), among these 9 women and 44 men. In terms of organisational level, 20 worked at a national/state level, 31 at a local level, and 2 in a private corporation/independent unit. Thirty-two had a civil educational background, 18 had a police educational background, and 3 had both. As per educational level, 1 had a Ph.D., 15 had an MSc, 34 had a BSc, and 3 had other education without a degree.

In terms of experience with criminal investigation within law enforcement, the overall distribution was: 0–1 years: 32%, 2–4 years: 25%, 5–9 years: 32%, 10–15 years: 5%, above 15 years: 5%. In terms of experience with DF within law enforcement the overall distribution was: 0–1 years: 28%, 2–4 years: 32%, 5–9 years: 32%, 10–15 years: 8%, and above 15 years: 0%.

The main processing/analysis tool used during the experiment was according to the participants: Magnet Axiom: 53%, X-Ways: 28% EnCase: 6%, Autopsy 6%, Forensic Toolkit (FTK) 4%, The Sleuth kit (TSK): 2%, Other: 2%.

To achieve a good distribution of the background variables across experimental conditions, participants were quasi-randomly assigned to groups (Shadish et al., 2002). The participants were first matched in groups of four with participants with similar educational background and level, organisational level, and experience within law enforcement. They were then randomly assigned to an experimental condition. The demographics per group are shown in Table 1.

2.2. Material

2.2.1. Evidence file

To generate data for the experiment, the participants were tasked with analysing an evidence file. We used the evidence file from Digital Corpora (Garfinkel et al., 2009). The file was approximately 3 GB, and in E01 format (nps-2008-jean.E01 and nps-2008-jean.E02). The evidence file had an NTFS file system and Microsoft Windows XP operating system, and the timestamps in the file were from 2008. The owner was defined as Jean User, and contained the typical programs and files that would be present on a work computer, such as programs for handling documents, spreadsheets or presentations, e-mail, chat, internet browsing, etc.

2.2.2. Scenarios

A scenario explaining an incident of confidential information leakage was provided to the participants (see Appendix 1). Participants in the experimental groups received additional contextual information strongly indicating guilt (referred to as the Strong Guilt group), ambiguous and weak indications of guilt (referred to as Weak Guilt group), and innocence (referred to as the Innocence group). Also, we had a control group that only received the scenario, and no additional contextual information (Control group).

2.3. Procedure

After signing and returning an information letter and a consent form, the participants were given a background survey (Part 1 survey). They were free to choose the day for completing the experiment. They were allowed to work at their normal

Table 1
Demographics – percentage* distribution of background variables per group.

Demographics		Control N = 16	Strong Guilt N = 12	Weak Guilt N = 13	Innocent N = 12
Country*	Norway	75	83	85	92
	Other	25	17	15	8
Gender*	Men N = 44	81	100	85	67
	Women N = 9	19	0	15	33
Educational background*	Civil	63	58	54	66
	Police/both	38	42	46	33
Educational level*	PhD	6	0	0	0
	MSc	25	33	23	33
	BSc	69	58	69	58
	Lower	0	8	8	8
Organisational level*	National	31	25	54	42
	Local	63	67	46	58
	Private industry	6	8	0	0
Mean years of experience within law enforcement	With crim. inv.	5	6	5	6
	With DF	5	4	3	4

workstation, and with the processing/analysis software of their own choice. They were instructed to work alone, and not to consult colleagues during the experiment.

The day before the experiment day, they were given the link for downloading the evidence file, and they were informed that they could process the file in advance, but not start on the analysis until they had received information and templates the next morning. On the experiment day, they received the scenario and templates for the log and analysis reports on e-mail.

The participants received different contextual information, but this was not visible from the document title – which all contained the title “READ THIS description of the experiment”. They were reminded that they should work alone, and not communicate with anyone about the experiment. They were asked to make notes in the log while analysing, and the template contained fields for actions, justification/purpose, result, whether the result was bookmarked, and assessment of the result. They were also asked to note the starting time and completion time in the log, and if they were interrupted, also the periods they worked on the experiment.

The participants were asked to reserve 4–5 h on the analysis and writing the report. The analysis report template contained the heading ‘Analysis report’, and a line for the name of the report author and date. They were instructed to bookmark files of interest and export them to a pdf and return this document by e-mail together with the analysis report and log when the experiment was completed. After receiving these documents from the participants, they were provided with the final survey (Part 2 survey). In this survey, they were provided with a list of information that was located on the evidence file (see Appendix 2, A1, A3–A11), and asked whether they identified the particular trace and if so, to assess whether they thought it indicated guilt, innocence or was ambiguous for the suspect Jean. They were also asked if they identified traces in addition to the list that indicated guilt/innocence/ambiguity for Jean, with the same alternatives as described above.

The output of the experiment included in this article was the following:

- Part 1 Survey: Background information
- Analysis reports
- Part 2 Survey: About how the analysis was performed and assessment of the findings

2.4. Analysis and quality control

The analysis was done in IBM SPSS Statistics Release 26.0.0.64

bit edition.

The observations of traces (see section 3.1.1 and Appendix 2) were coded in dichotomous variables “found” and “not found”. The interpretations of observed traces (see section 3.1.2 and Appendix 2) were coded in dichotomous variables “corresponding interpretation” and “non-corresponding interpretation”. The conclusions (see section 3.1.3 and Appendix 3) were coded as “not found”, “guilt”, “innocence” or “ambiguous”.

Kruskal–Wallis test was used for the analysis of biasability (the difference between groups). Kruskal–Wallis is a non-parametric rank-based test used for small sample sizes which do not meet the assumption of a normal distribution and tests whether there is a significant difference between groups (Skovlund and Fenstad, 2001). The reported results are the H statistic, degrees of freedom, and the p-value; e.g. $H(3) = 9.438, p = .024$. Mean ranks are the average of the ranks for all observations within each sample, and are used for calculating the statistic (H). Mean ranks should be interpreted as when a group’s mean rank is higher than the overall average rank, the observation values in that group tend to be higher than those of the other groups.

Krippendorff’s Alpha coefficient was used for the analysis of reliability (the variation within groups) and was chosen as measurement since it is a generalisation of several known reliability indices (Krippendorff, 2011). It and enables the judgement of a variety of data with the same reliability standard, and applies to any number of observers, any number of categories, scale value or measures, any level of measurement, large or small sample sizes, and datasets with incomplete or missing data (Krippendorff, 2011). The method can also be used for where multiple coders are coding data (Nili et al., 2017). Values of $\alpha = 0.80$ or higher are considered a strong level of consistency, while values lower than $\alpha = .667$ are considered an inadequate level of consistency (Hayes and Krippendorff, 2007).

Krippendorff’s Alpha coefficient is often used to assess agreement among coders in qualitative analysis and was used for the assessment of inter-coder agreement in the current study. The background information from Part 1 and assessment of findings from Part 2 (Section 3.1.3) was multiple-choice and required only minimal interpretation. The observations of traces (A) (Section 3.1.1) and interpretations of observed traces (B) (Section 3.1.2), required some interpretation and were coded as dichotomous variables (values: identified – not identified). To assess inter-coder agreement, 10% of the reports (A and B) were coded by a person who was not engaged with the study. Krippendorff’s Alpha coefficient (Hayes and Krippendorff, 2007) was computed, with the result of $\alpha = 0.91$, which is considered a strong level of agreement

(Hayes and Krippendorff, 2007).

3. Results and discussion of biasability and reliability

3.1. Biasability in DF examinations

The first research question was to explore whether DF examiners were biased by contextual information in their observations, interpretations of observations, or conclusions. We gave the same evidence file and baseline scenario to the participants, but different contextual information (indicating strong guilt, weak guilt, or innocence – see Appendix 1). A control group received only the baseline scenario and no contextual information. The results with regards to observations are presented in section 3.1.1, interpretations of observations in section 3.1.2, and conclusion in section 3.1.3. The results concerning biasability are discussed in section 3.2.

3.1.1. Biasability in observations of evidence (A)

The evidence file provided to the DF examiners was relatively small (3 GB), but there was a lot of information to explore. The DF examiners documented relevant findings in the analysis report, and 11 different traces were selected for comparison of the proportion of observations. Some of the traces were easier to find on the evidence file, such as e-mail content, documents, and chat. Other traces would require more in-depth examination, such as e-mail headers or discovering the mounting of USB. The traces A1-A11 are not very complex, and finding these traces would require basic DF skills.

None of the participants observed all 11 traces (A). Overall, fourteen (26%) of the participants found 1–4 traces, 35 (66%) found 5–8 traces and 4 (8%) found 8–10 traces. When comparing the groups, higher numbers were observed in the Guilt groups, then the Control group, and the lowest in the Innocence group. Specifically, the Weak Guilt group had the highest number of observed traces per participant ($M = 6.9, SD = 2.0$, range 4–10), followed by the Strong Guilt group ($M = 5.7, SD = 1.2$, range 3–7), the Control group ($M = 5.4, SD = 1.7$, range 3–9) and the Innocence group ($M = 4.5, SD = 1.8$, range 1–7).

Table 2 shows the proportion of participants per group that observed the trace, and a detailed description of the traces is provided in Appendix 2. The mean and SD refer to the proportions of observed traces within each of the groups. The possible number of observations was 145.75 per group (11 observations x 53 participants/4 groups = 145.75). To compare the number of observations

between groups, they were adjusted for group size (Number of observations x .63 (Control group), x .83 (Strong Guilt group), x .77 (Weak Guilt group), x .83 (Innocence group).

The Innocence group observed the least number of traces, indicating that they were biased to find less evidence. The Guilt groups had the highest number of traces, indicating that they were biased to find more evidence. The Control group was between the Innocence and Guilt groups. However, there was very little difference between the Strong Guilt group and the Control group in the proportion of observations, which indicates that the Strong Guilt context (the suspect had confessed) did not bias the participants to observe more traces than an examination without such context. The Weak Guilt group observed significantly more traces, suggesting that the ambiguous Weak guilt context (wage conflict where the suspect had taken side with the workers) biased the Weak Guilt group to find more traces. These differences were statistically significant ($p < .05$) as reflected by a Kruskal–Wallis test between groups: $H(3) = 9.438, p = .024$, with mean ranks of 36.62 for the Weak Guilt group, 28.08 for the Strong Guilt group, 24.84 for the Control group, and 18.38 for the Innocence group.

Summarised, a statistically significant effect from contextual information was observed on the proportion of observations. The Weak Guilt context led to most observations, and Innocence context led to fewer observations. The Strong guilt context did not have a significant effect on observations compared to the Control group.

3.1.2. Biasability in interpretations of observed evidence (B)

During an examination of digital evidence, DF examiners use their expertise to interpret and derive meaning from the observed traces. These interpretations are presented in analysis reports as various forms of opinion statements, which in this paper are referred to as interpretations of observed traces, or just interpretations. Examples from the current study were e.g. “m57plan.xls was not available in the image file” or “the two spreadsheets had similar information but different file names”. These are often not the final conclusion, but would often underpin or justify the final conclusion. Seven different interpretations of observed traces (B) were selected for comparison and verified by a DF expert with no role in the study. If the information was missing, incorrect, or incomplete, it was coded as a “non-corresponding interpretation”. A detailed description of the criteria for being coded as a corresponding interpretation is provided in Appendix 2.

The possible number of interpretations was 92.75 per group (7 interpretations x 53 participants/4 groups = 92.75). To compare the

Table 2
Proportion of DF examiners per group that observed the traces (A).

Observations of traces	Control (N = 16)	Strong Guilt (N = 12)	Weak Guilt (N = 13)	Innocence (N = 12)	Total (N = 53)	Range
(A1)	.69	1.00	.85	.50	.76	.50
(A2)	.94	1.00	.92	.92	.94	.08
(A3)	.19	.33	.23	.08	.21	.25
(A4)	1.00	.92	1.00	.83	.94	.17
(A5)	.44	.58	.62	.25	.47	.37
(A6)	.88	.92	1.00	.92	.93	.12
(A7)	.38	.08	.54	.00	.26	.54
(A8)	.19	.00	.39	.00	.15	.39
(A9)	.19	.08	.31	.08	.17	.23
(A10)	.25	.33	.39	.42	.34	.17
(A11)	.25	.42	.69	.50	.45	.44
Mean	.49	.52	.63	.41	.51	
SD	.32	.39	.28	.36	.32	
Total observations	86	68	90	54		
Observations adjusted for group size	54	56	69	45		

Table 3
Proportion of DF examiners per group that interpreted the traces correctly (B).

Interpretations of observed traces	Control (N = 16)	Strong Guilt (N = 12)	Weak Guilt (N = 13)	Innocence (N = 12)	Total (N = 53)	Range
(B1)	.19	.42	.08	.33	.26	.34
(B2)	.19	.33	.23	.0	.19	.33
(B3)	.38	.42	.31	.25	.34	.17
(B4)	.88	.92	1.00	.92	.93	.12
(B5)	.25	.17	.15	.50	.27	.35
(B6)	.19	.08	.31	.08	.17	.23
(B7)	.25	.33	.46	.33	.34	.21
Mean	.33	.38	.36	.34	.36	
SD	.25	.27	.31	.30	.26	
Total interpretations	37	32	33	29		
Interpretations adjusted for group size	23	27	27	22		

number of observations between groups, they were adjusted for group size (Number of interpretations x .63 (Control group), x .83 (Strong Guilt group), x .77 (Weak Guilt group), x .83 (Innocence group), see Table 3 for details. The groups were comparable and the variation between them was quite small. A Kruskal–Wallis test performed at the total of observations showed no significant effect ($p < .05$) from contextual information between groups: $H(3) = 0.502, p = .918$.

3.1.3. Biasability in conclusions about guilt/innocence/ambiguity (HEP 8)

After writing the report from the analysis and submitting the result, the participants were provided with a survey about traces (A) from the experiment. They were asked to assess whether they identified the trace, and if so, whether it indicated guilt, innocence, or was ambiguous for the suspect. Traces that was identified by a minimum of 35% of the DF examiners were included here: A1, A3–A11, see detailed description in Appendix 2 and 3.

Of a total of 529 ratings of the traces, 110 were rated as “not found”, and 419 were rated as indications of guilt, innocence or ambiguous. Among these 419 ratings, there were 22 ratings of guilt (Control group = 5, Strong Guilt group = 9, Innocence group = 3,

Weak Guilt group = 5), 125 ratings of innocence (Control group = 46, Strong guilt group = 26, Innocence group = 29, Weak Guilt group = 24), and 272 ratings of ambiguous (Control group = 81, Strong Guilt group = 65, Innocence group = 56, Weak Guilt group = 70). The proportion of “not found” among the groups was: Control group = 28 (17.5%), Strong Guilt group = 20 (16.7%), innocence group = 33 (27.3%), Weak Guilt group = 29 (22.7%).

Fig. 2 shows the percentage of ratings of guilt, innocence and ambiguous within those who found the evidence, and the value “not found” is excluded. A Kruskal–Wallis test performed at the total of conclusions showed no statistically significant effect ($p < .05$) for contextual information between groups, Not found: $H(3) = 3.182, p = .364$, Guilt: $H(3) = 3.357, p = .316$, Innocence: $H(3) = 3.261, p = .353$, Ambiguous: $H(3) = 0.914, p = .822$.

3.2. Discussion of biasability

The research question related to biasability was: *Are DF examiners biased by contextual information when they make observations, interpretations of observations, or in their conclusions during the analysis of digital traces?*

The results show that the observations were biased by the

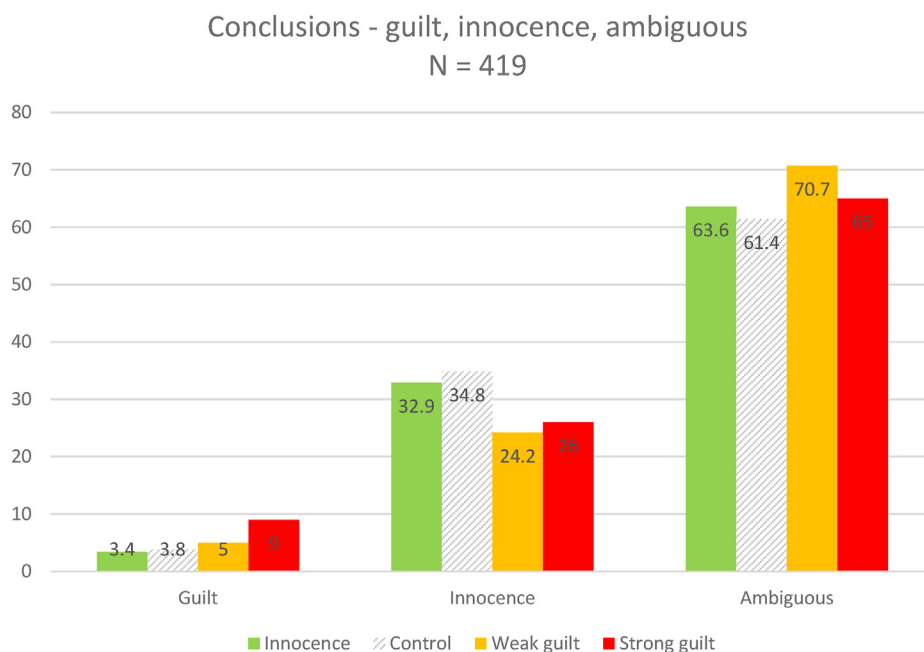


Fig. 2. Proportion of guilt, innocence, and ambiguous ratings for 10 traces by the groups.

contextual information, where more traces were observed when Guilt contexts were introduced, and fewer traces were observed when Innocence context was introduced. The examination of biasability at the interpretation and conclusion level did not provide statistically significant results.

The proportion of observations was obtained from the participants' analysis reports. The traces that were not reported may have not been observed or may have been observed and overlooked or explained away as not relevant.

The results suggest that a DF examiner who believes that the suspect is innocent, observes fewer traces, and thus may have less information for developing explanations about what may have happened, including scenarios involving that the suspect may have contributed to or committed the crime. If the traces are overlooked or explained away due to an innocence bias, a possible consequence may be that important traces are not reported, or might be described and framed in a way that fits with the innocence hypothesis. The following is an illustrative example of such an 'innocence framing' of the findings in the conclusion: "It is very likely that M57. biz was exposed to a phishing attack and that the police hypothesis about Jean being innocent, and rather was framed, is correct" (P-31 – translation from Norwegian).

In the same manner, the results suggest that a DF examiner who believes that the suspect may be guilty, observe more traces, and thus obtain more information basis. The following is an example of 'guilt framing' of the findings in the conclusion: "The employees involved are Alison SMITH, manager of M57. biz and Jean JONES, CFO of M57. biz. Additionally, e-mail communication would also suggest that tuckgorge@gmail.com also has an involvement in the leakage of confidential information" (P-62).

The proportion of observed traces according to the survey (Appendix 3) is higher than what was the proportion that was documented in the analysis reports (Table 2). This difference indicates that the traces may have been overlooked or explained away rather than that they actually were not discovered, and sheds light on how the DF examiners contribute to the construction of the criminal case by including and excluding information in their analysis reports.

The DF examiners were allowed to use processing/analysis tools of their own choice. Processing tools may display the information differently. Therefore, we explored whether there was a relationship between the processing tool and the observations, but no relationship was found. The findings of biased observations uncovered in this study are probably not related to what the processing tool displayed or not, but rather about what the DF examiner looked for and perceived as relevant.

A covariate that may have played a role in the observations, interpretations, or conclusions is the data itself (see Dror, 2020). The evidence file contains a lot of "rich" information that may have impacted the DF examiners during the analysis, such as e-mail and instant messaging communication.

Another critical point, often overlooked, is how bias may impact the results by influencing testing strategies (Dror, 2020). An important aspect of this, demonstrated in our study, is how bias may influence whether the examiner ends or continues the further examination. This has been observed in other forensic domains (Dror, 2020), but not in DF. As detailed above, the data showed that when Guilt contexts were received, and particularly the ambiguous Weak Guilt context, the DF examiners continued to look for and observe more traces, in contrast to when Innocence context was received, where the examiners stopped looking for traces much

earlier and hence observed fewer traces.

The findings concerning biasability at the observation level shed light on the point at which bias starts affecting the outcome. This knowledge is critical for designing effective bias mitigation measures for DF work, intending to minimize human error.

3.3. Reliability in DF examinations

The second research question was to explore whether DF examiners were consistent when they make observations, interpretations of observations, or conclusions during the analysis of digital traces.

A calculation of Krippendorff's Alpha coefficient (Hayes and Krippendorff, 2007) was performed to measure the consistency within each group, where the participants had analysed the same evidence file based on the same contextual information. According to Hayes and Krippendorff (2007), values of $\alpha = 0.80$ or higher are considered a strong level of consistency, while values lower than $\alpha = 0.667$ are considered an inadequate level of consistency.

For observation and conclusions of traces (A) and interpretations of observed traces (B), the results suggest overall low/inadequate reliability ($\alpha < 0.667$) (see Table 4). The highest reliability score was found at the observation level in observation of traces (A) for examiners receiving Strong Guilt context ($\alpha = 0.51$) and Innocence context ($\alpha = 0.44$).

The most extreme within-group variation at the conclusion level is the variation from rating a trace as an indication of guilt vs an indication of innocence (see Appendix 3). This variation was observed for all groups, and highest for Strong Guilt (40%, A3, A4, A5, and A6) followed by Control (30%, A1, A4, and A6). For Weak Guilt and Innocence, the variation between guilt and innocence ratings was observed in 20% of the evidence (A1 and A6).

3.4. Discussion of reliability

The research question related to reliability was: *Are DF examiners consistent with one another when they make observations, interpretations of observations, or conclusions during the analysis of digital traces?*

The results suggest what is considered a low/inadequate ($\alpha < 0.667$) consistency between DF examiners who examined the same evidence file based on the same contextual information. The low consistency was found on all the examined levels: observations of traces, interpretations of observed traces, and conclusions. The results suggest that if a DF examiner performs an analysis of an evidence file, and another DF examiner would do a re-analysis of the same evidence file, the chances of reaching consistent results are low. This finding is of great importance, since it indicates, quite contrary to the perception of digital evidence as objective and credible, that digital evidence may be affected by the human examiner in all the stages we have examined – during the observation, the interpretation of observation and when forming a conclusion. The low reliability would thus not only have implications for reporting opinion evidence but also for factual/technical reporting.

Although high reliability between DF examiners is anticipated, it is important to be aware that consistency does not imply accuracy or validity. Consistency may arise from a variety of reasons, for example when examiners are consistently biased by similar biasing contextual information. If the irrelevant contextual information biases the DF examiners in the wrong direction, they may perform

Table 4Krippendorff's α coefficient of reliability for observations of traces, interpretations of observed traces, and conclusions.

Reliability	Control N = 16 N = 132*	Strong Guilt N = 12 N = 100*	Weak Guilt N = 13 N = 99*	Innocence N = 12 N = 88*
Observations (A)	.35	.51	.26	.44
Interpretations of observations (B)	.20	.20	.30	.30
Conclusions (A)*	.26	.22	.20	.30

similar and consistent observations, and all reach the same incorrect conclusions. In such a case, the consensus is reached for the wrong reasons. Consistency is thus no proof of validity.

This entails that quality measures should not only be directed towards the tools and technology, but also the human. It is not possible to calibrate a human the same way as a technical instrument, but measures such as blind proficiency testing through the use of fake test cases may provide knowledge about human performance (Garrett, 2021). Quality control measures such as verification reviews and re-examination of the evidence (Horsman and Sunde, 2020; Sunde and Horsman, 2020) may be important measures to detect possible erroneous or misleading results.

4. General discussion

To the best of our knowledge, this is the first study to explore reliability and biasability in DF decision making. The paper seeks to advance the current knowledge about the potential sources of error in digital forensic investigations. The study aimed at examining whether contextual information biases the DF examiners' observations, interpretations of observations, and conclusions during the examination of a dataset, and to further explore whether DF examiners were consistent in their observations, interpretations of observations, and conclusions when they analysed the same dataset with the same contextual information.

The results showing low reliability are probably the most important findings in this study. The study shows low reliability between experts examining the same evidence file with the same contextual information, in their observations, interpretation of observations, and their conclusions. The results thus indicate a low probability for a second DF examiner observing and interpreting the same traces, and concluding consistently compared to the initial DF examiner (given the same evidence file and contextual information). The study contributes to a growing body of knowledge about the lack of reliability in forensic decision making (e.g. Ulery et al., 2012; Lidén and Dror, 2020), and highlights the need for systematic implementation of quality control measures such as blind peer review (Horsman and Sunde, 2020; Sunde and Horsman, 2020) in DF examinations.

In terms of biasability, the pre-study suggests that when DF work is commissioned, it is common to provide contextual information to the DF examiner and that some of this information may be task-irrelevant. The current study showed that DF examiner observations were affected by the biasing contextual information.

Our results are consistent with basic research on confirmation bias (Nickerson, 1998). The current study contributes to the growing volume of studies indicating that forensic science decision making can be affected by contextual information (e.g., Elaad et al., 1994; Cooper and Meterko, 2019) and that confessions may affect the collection and evaluation of evidence (Hasel and Kassin, 2009; Kassin et al., 2012; Kukucka and Kassin, 2014).

The findings regarding biasability and reliability suggest that human factors have a significant impact on the outcome of a DF examination and that the results of DF examinations should not be presumed to be objective and credible. It is thus necessary to gain more knowledge on how to minimize and/or detect bias in DF decision making.

To minimize bias during forensic work, excluding or blinding of irrelevant contextual information has been identified as a possible solution for several forensic science domains. This would also be an important measure in DF examination, however, most likely it would not totally solve the contextual bias challenge. This is due to the nature of digital evidence, which normally contains lots of both task-relevant and irrelevant information within the dataset under investigation which cannot be easily excluded. A bias countermeasure could be that the starting point of every DF examination is a balanced set of alternative hypotheses, where both innocence and guilt hypotheses are represented, that the DF examiner systematically considers all the hypotheses during the analysis, and report the results accordingly. Transparency about what they knew (what task-relevant and task-irrelevant information they received), and which hypotheses guided the examination are of great importance to facilitate a review of the process and result, and enable detection of biased decision making.

In terms of ecological validity, the experimental setting was designed to be as close as possible to the real work situation of the DF examiner. The pre-study of submission forms indicated that it is common to provide contextual information when DF work is commissioned and that some of this information may be task-irrelevant. The contextual information provided to the DF examiners in the current study would thus be within what would be a normal situation encountered in actual casework. The DF examiners were allowed to use their normal workstation, and the analysis software of their preference, and could report the results as they would do in their casework.

There were some limitations in the experimental setting that should be commented on. A possible limitation of the study was the sample size of 53 DF examiners, which may limit the statistical power. The time frame may also have been a limitation, and a few commented that 4–5 h may have been less time than they normally would have spent on the analysis of an evidence file. However, the evidence that was observed (A) or interpreted (B) by at least 31% of a group was included in the further analysis of biasability and reliability. Since the study –as most studies– used volunteer participation, there is also a risk of self-selection bias. Since they knew they were taking part in research, a Hawthorne-effect (e.g. Leedy and Omerod, 2014) may have occurred, however – we suggest this would rather impact the results in a “positive direction” in terms of evidence observation and interpretation of the observed evidence, for example, that they examined the evidence file more thoroughly and accurately than they would normally do, and thus observed more traces.

In terms of future directions for research, it is important to explore how bias could be minimized during DF work. To minimize potentially flawed results from DF examinations, there seems to be a need for quality assurance and control, which also calls for more research on effective measures (see e.g. [Mattijssen et al., 2020](#); [Sunde and Horsman, 2020](#)). The results related to biasability underpins a need for research on contextual bias and context management. There is a need to gain more knowledge about what information has the most biasing effects, and which contextual information is task-relevant vs task-irrelevant. With this knowledge, it is possible to design context management procedures for DF that ensure access to the task-relevant contextual information, while excluding or limiting access to task-irrelevant contextual information.

5. Conclusion

The study explored the reliability and biasability in DF decision making and involved 53 DF examiners who analysed the same evidence file. The aim was to examine whether contextual information biases the DF examiners' observations, interpretations, or conclusions and whether DF examiners make consistent observations, interpretations, or conclusions when they examine the same evidence file with the same contextual information.

The results showed that the DF examiners' observations were affected by the biasing contextual information. To minimize bias, it is important to ensure that DF examinations are based on task-relevant contextual information, and minimize the exposure to task-irrelevant contextual information. The results also showed low reliability between DF examiners in their observations, interpretations of observations, and conclusions. The low reliability indicates that there is a serious and urgent need for quality assurance in DF examinations and control of DF results to prevent erroneous results from cascading into the investigation process.

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.

Acknowledgements

We thank the DF examiners who dedicated some of their valuable time and participated in this study, and also Torstein Schjerven and Jørn Helge Jahren for their contribution to pilot testing of the quasi-experiment. We are also very grateful for the support from Håvard Aanes on the statistical measurements used in this study and for the contribution to the control of inter-coder reliability from Christine S. Nordsletten and Torstein Eidet. We thank Helene O. I. Gundhus from the University of Oslo, and Fergus T. Toolan, Johanne Yttri Dahl, and Gunnar Thomassen from the Norwegian Police University College for insights and constructive criticism of former versions of the paper. We also thank the anonymous reviewers for their constructive critique and helpful comments which have improved the paper.

This paper was funded by a grant from the Norwegian Police University College awarded to the first author for pursuing a PhD.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fsidi.2021.301175>.

APPENDIX 1. Case vignettes

Confidential information leakage

M57. biz is a small US based company, with office in your country. The company, which develops and sells body art equipment (tattoo, piercing etc.), is in the start-up phase. The manager for the M57. biz office in your country is Alison Smith, and the CFO is Jean Jones. The company has 4 programmers, 2 in marketing, and 1 in business development. Only Alison and Jean have a permanent office space, while the other employees work from home office. All employees participate in a daily online meeting. There are in-person meetings for all employees in the M57. biz office once every two weeks. Most documents are exchanged by e-mail.

A spreadsheet (m57plan.xls) containing confidential information was recently posted as an attachment in a forum of a competitor's website. When this was discovered, Alison reported the incident to the police as information theft. Alison told police that Jean, the CFO, was responsible for updating the spreadsheet, and that it was probably sent from Jean's computer.

The attachment posted on the competitor's website looked like this:

M57.biz company				
Name		Position	Salary	SSN (for background check)
Alison	Smith	President	\$140,000	103-44-3134
Jean	Jones	CFO	\$120,000	432-34-6432
Programmers:				
Bob	Blackman	Apps 1	90,000	493-46-3329
Carol	Canfred	Apps 2	110,000	894-33-4560
Dave	Daubert	Q&A	67,000	331-95-1020
Emmy	Arlington	Entry Level	57,000	404-98-4079
Marketing				
Gina	Tangers	Creative 1	80,000	980-97-3311
Harris	Jenkins	G & C	105,000	887-33-5532
BizDev				
Indy	Counterchng	Outreach	240,000	123-45-6789
Annual Salaries			\$1,009,000	
Benefits			30%	\$302,700

You are tasked with analysing a copy of the hard drive from Jeans computer, and find out:

What has happened, and what was Jeans involvement in the reported incident?

Participants other than the Control Group received one of the following additional narratives:

(Strong Guilt):

Jean was arrested for information theft, and in a police interview earlier today, she confessed that she had committed the criminal offense.

(Weak Guilt):

Alison told police that there has been a long-lasting wage dispute with the programmers in the firm, who claim to be underpaid. One of the programmers - Emmy Tuckford Arlington – has fronted the dispute on behalf of the programmers in M57. biz. Jean

has supported the programmers in this conflict, and has told Alison that the company can afford to pay them better salaries. Jean is about to be interviewed by the police about the reported incident. However, the chief investigating officer wants an analysis of Jean's computer before the police interview, to look for traces indicating that she was involved in the reported incident.

(Innocence):

As a result, Jean was arrested for information theft and questioned about the incident in a police interview. However, after the police interview, the police believe she is innocent, and that she was framed during a phishing attack.

APPENDIX 2. Observed traces and interpretations of traces

Observed traces (A) (coding: found/not found)

1. Spreadsheet with the filename m57biz.xls stored in the Desktop folder. It contained information similar to what was leaked.
2. Spreadsheet with the filename m57biz.xls located in the outlook.pst file. It contained information similar to what was leaked.
3. AIM chat between alison57 and Jean User. There are indications of that someone might be impersonating Alison – the boss is referred to as a male.
4. Mail correspondence between jean@m57.biz and Alison@m57.biz. The address is used by the boss – Alison, which asks Jean to create a spreadsheet with information about the employees.
5. Mail correspondence between Jean@m57.biz and Alex@m57.biz. The address is also used by the boss – Alison, with information relevant to the information leakage.
6. Mail correspondence between Jean@m57.biz and Tuckgorge@gmail.com, outside the firm M57. biz e-mail domain. Tuckgorge@gmail.com is used together with the display name Alison@m57.biz, and is recipient of the mail attachment m57biz.xls.
7. Mail correspondence between Jean@m57.biz and Bob@m57.biz, who asks about the leaked information.
8. Mail correspondence between Jean@m57.biz and Carol@m57.biz, who comments that she have noticed that something is wrong.
9. Trace of mounting of a USB (type, model), which may contain relevant traces or information about the incident.
10. Several user accounts had been activated, including the Administrator account.
11. Several user accounts had been activated, including the Devon user account.

Valid interpretations of observed traces (B) (coding: corresponding interpretation/non-corresponding interpretation)

1. Excel spreadsheet m57plan.xls was not on the evidence file.
2. That excel spreadsheets m57plan.xls and m57biz.xls were different/did not contain the exact same information.
3. That excel spreadsheets m57biz.xls located in Desktop folder and in outlook. pst had the same hash value.
4. That the mail with the attachment m57biz.xls was sent to tuckgorge@gmail.com and not alison@m57.biz (which was only the display name in e-mail that was replied to).
5. That simsong@xy.dreamhostps.com was defined as return path in the header of the e-mail message where the spreadsheet with

confidential information was requested. This e-mail message was replied to and the spreadsheet m57biz.xls was enclosed.

6. That the USB was mounted shortly before the m57biz.xls was sent out of M57. biz domain.

7. That other user accounts (Devon and/or Administrator) than the Jean user account were active around the time when the m57biz.xls was sent.

APPENDIX 3. Conclusions of whether the trace indicated guilt/innocence/ambiguous

(Coding: not found, guilt, innocence, ambiguous)

Trace	N = Control (16) Strong guilt (12) Innocence (12) Weak guilt (13)	% of N stated that they identified the trace	Rating: Guilt*	Rating: Innocence*	Rating: Ambiguous*
A1	Control Strong guilt Innocence Weak guilt Mean	100% 100% 92% 100% 98%	2 (13%) 5 (42%) 1 (9%) 3 (23%)	1 (6%) 0 (0%) 2 (18%) 1 (8%)	13 (81%) 7 (58%) 8 (73%) 9 (69%)
A3	Control Strong guilt Innocence Weak guilt Mean	81% 67% 67% 54% 67%	0 (0%) 1 (13%) 0 (0%) 0 (0%)	3 (23%) 1 (13%) 1 (13%) 0 (0%)	10 (77%) 6 (75%) 7 (88%) 7 (100%)
A4	Control Strong guilt Innocence Weak guilt Mean	94% 100% 83% 100% 94%	1 (7%) 1 (8%) 0 (0%) 0 (0%)	7 (47%) 7 (58%) 6 (60%) 6 (46%)	7 (47%) 4 (33%) 4 (40%) 7 (54%)
A5	Control Strong guilt Innocence Weak guilt Mean	87% 100% 83% 100% 93%	0 (0%) 1 (8%) 0 (0%) 0 (0%)	6 (43%) 5 (42%) 4 (40%) 3 (23%)	8 (57%) 6 (50%) 6 (60%) 10 (77%)
A6	Control Strong guilt Innocence Weak guilt Mean	100% 100% 92% 100% 98%	2 (13%) 1 (8%) 1 (9%) 2 (15%)	13 (81%) 8 (67%) 10 (91%) 8 (62%)	1 (6%) 3 (25%) 0 (0%) 3 (23%)
A7	Control Strong guilt Innocence Weak guilt Mean	94% 100% 83% 92% 92%	0 (0%) 0 (0%) 1 (10%) 0 (0%)	9 (60%) 3 (25%) 4 (40%) 2 (17%)	6 (40%) 9 (75%) 5 (50%) 10 (83%)
A8	Control Strong guilt Innocence Weak guilt Mean	87% 83% 67% 85% 81%	0 (0%) 0 (0%) 0 (0%) 0 (0%)	6 (43%) 2 (20%) 2 (25%) 4 (36%)	8 (57%) 8 (80%) 6 (75%) 7 (64%)
A9	Control Strong guilt Innocence Weak guilt Mean	62% 50% 50% 46% 52%	0 (0%) 0 (0%) 0 (0%) 0 (0%)	1 (10%) 0 (0%) 0 (0%) 0 (0%)	9 (90%) 6 (100%) 6 (100%) 6 (100%)
A10	Control Strong guilt Innocence Weak guilt Mean	56% 67% 54% 38% 54%	0 (0%) 0 (0%) 0 (0%) 0 (0%)	0 (0%) 0 (0%) 0 (0%) 0 (0%)	9 (100%) 8 (100%) 6 (100%) 5 (100%)
A11	Control Strong guilt Innocence Weak guilt Mean	62% 67% 50% 61% 60%	0 (0%) 0 (0%) 0 (0%) 0 (0%)	0 (0%) 0 (0%) 0 (0%) 0 (0%)	10 (100%) 8 (100%) 6 (100%) 8 (100%)

*% of those who found the trace.

References

- Casey, E., 2002. Error, uncertainty and loss in digital evidence. *International Journal of Digital Evidence* 1 (2).
- Christensen, A.M., Crowder, C.M., Ousley, S.D., Houck, M.M., 2014. Error and its meaning in forensic science. *J. Forensic Sci.* 59 (1), 123–126. <https://doi.org/10.1111/1556-4029.12275>.
- Cooper, G.S., Meterko, V., 2019. Cognitive bias research in forensic science: a systematic review. *Forensic Sci. Int.* 297, 35–46. doi:10.1016/j.forsciint.2019.01.016.
- Dror, I.E., 2016. A hierarchy of expert performance. *Journal of Applied Research in Memory and Cognition* 5 (2), 121–127. <https://doi.org/10.1016/j.jarmac.2016.03.001>.
- Dror, I.E., 2020. Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias. *Anal. Chem.* 92 (12), 7998–8004. <https://doi.org/10.1021/acs.analchem.0c00704>.
- Dror, I.E., Hampikian, G., 2011. Subjectivity and bias in forensic DNA mixture interpretation. *Sci. Justice* 51 (4), 204–208. <https://doi.org/10.1016/j.scijus.2011.08.004>.
- Dror, I.E., Murrice, D., 2018. A Hierarchy of Expert Performance (HEP) applied to forensic psychological assessments. *Psychol. Publ. Pol. Law* 24 (1), 11–23. <https://doi.org/10.1037/law0000140>.
- Dror, I.E., Charlton, D., Péron, A.E., 2006. Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Sci. Int.* 156 (1), 74–78. <https://doi.org/10.1016/j.forsciint.2005.10.017>.
- Elaad, E., Ginton, A., Ben-Shakhar, G., 1994. The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *J. Behav. Decis. Making* 7 (4), 279–292. <https://doi.org/10.1002/bdm.3960070405>.
- Gardner, B.O., Kelley, S., Murrice, D.C., Blaisdell, K.N., 2019. Do evidence submission forms expose latent print examiners to task-irrelevant information? *Forensic Sci. Int.* 297, 236–242. <https://doi.org/10.1016/j.forsciint.2019.01.048>.
- Garfinkel, S., Farrell, P., Roussev, V., Dinolt, G., 2009. Bringing Science to Digital Forensics with Standardized Forensic Corpora, DFRWS 2009, Montreal, Canada. Available at: <https://simson.net/clips/academic/2009.DFRWS.Corpora.pdf>. (Accessed 20 April 2020).
- Garrett, B.L., 2021. *Autopsy of a Crime Lab*. University of California Press.
- Garrett, B.L., Neufeld, P.J., 2009. Invalid forensic science testimony and wrongful convictions. *Va. Law Rev.* 95 (1), 1–97.
- Hartley, S., Winburn, A.P., 2021. A Hierarchy of Expert Performance (HEP) applied to forensic anthropology. *J. Forensic Sci.* (in press).
- Hasel, L.E., Kassir, S.M., 2009. On the presumption of evidentiary independence: can confessions corrupt eyewitness identifications? *Psychol. Sci.* 20 (1), 122–126. <https://doi.org/10.1111/j.1467-9280.2008.02262.x>.
- Hayes, A.F., Krippendorff, K., 2007. Answering the call for a standard reliability measure for coding data. *Commun. Methods Meas.* 1 (1), 77–89. <https://doi.org/10.1080/19312450709336664>.
- Horsman, G., Sunde, N., 2020. Part 1: the Need for peer review in digital forensics. *Forensic Sci. Int.: Digit. Invest.* 35, 301062. <https://doi.org/10.1016/j.fsidi.2020.301062>.
- Huang, C., Bull, R., 2020. Applying Hierarchy of Expert Performance (HEP) to investigative interview evaluation: strengths, challenges and future directions. *Psychiatr. Psychol. Law.* <https://doi.org/10.1080/13218719.2020.1770634>.
- James, J.L., Gladyshev, P., 2013. A survey of digital forensic investigator decision processes and measurement of decisions based on enhanced preview. *Digit. Invest.* 10 (2), 148–157. doi:10.1016/j.diin.2013.04.005.
- Kassin, S.M., Bogart, D., Kerner, J., 2012. Confessions that corrupt: evidence from the DNA exoneration case files. *Psychol. Sci.* 23 (1), 41–45. <https://doi.org/10.1177/0956797611422918>.
- Krippendorff, K., 2011. Computing Krippendorff's Alpha-Reliability. http://repository.upenn.edu/asc_papers/43.
- Kukucka, J., Kassin, S.M., 2014. Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias. *Law Hum. Behav.* 38 (3), 256–270. <https://doi.org/10.1037/lhb0000066>.
- Leedy, P.D., Ormrod, J.E., 2014. *Practical Research: Planning and Design*. Pearson Education.
- Lidén, M., Dror, I.E., 2020. Expert reliability in legal proceedings: “Eeny, meeny, miny, moe, with which expert should we go?”. *Sci. Justice.* <https://doi.org/10.1016/j.scijus.2020.09.006>.
- Mattijssen, E.J., Witteman, C.L., Berger, C.E., Brand, N.W., Stoel, R.D., 2020. Validity and reliability of forensic firearm examiners. *Forensic Sci. Int.* 307, 110112. <https://doi.org/10.1016/j.forsciint.2019.110112>.
- Nickerson, R.S., 1998. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* 2 (2), 175–220.
- Nili, A., Tate, M., Barros, A., 2017. A critical analysis of inter-coder reliability methods in information systems research. In: *Australasian Conference on Information Systems*. Hobart, Australia.
- Page, H., Horsman, G., Sarna, A., Foster, J., 2019. A review of quality procedures in the UK forensic sciences: what can the field of digital forensics learn? *Sci. Justice* 59 (1), 83–92. <https://doi.org/10.1016/j.scijus.2018.09.006>.
- Pollitt, M., Casey, E., Jaquet-Chiffelle, D.O., Gladyshev, P., 2018. A Framework for Harmonizing Forensic Science Practices and Digital/multimedia Evidence. The Organization of Scientific Area Committees for Forensic Science (OSAC), USA.
- Shadish, W.R., Cook, T.D., Campbell, D.T., 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth Cengage Learning.
- Skovlund, E., Fenstad, G.U., 2001. Should we always choose a nonparametric test when comparing two apparently nonnormal distributions? *J. Clin. Epidemiol.* 54 (1), 86–92. [https://doi.org/10.1016/s0895-4356\(00\)00264-x](https://doi.org/10.1016/s0895-4356(00)00264-x).
- Sunde, N., Dror, I.E., 2019. Cognitive and human factors in digital forensics: problems, challenges, and the way forward. *Digit. Invest.* 29, 101–108. <https://doi.org/10.1016/j.diin.2019.03.011>.
- Sunde, N., Horsman, G., 2020. Part 2: the phase-oriented advice and review structure (PARS) for digital forensic investigations. *Forensic Sci. Int.: Digit. Invest.* 36, 301074. <https://doi.org/10.1016/j.fsidi.2020.301074>.
- Tart, M., Pope, S., Baldwin, D., Bird, R., 2019. Cell site analysis: roles and interpretation. *Sci. Justice* 59 (5), 558–564. <https://doi.org/10.1016/j.scijus.2019.06.005>.
- Ulery, B.T., Hicklin, R.A., Buscaglia, J., Roberts, M.A., 2012. Repeatability and reproducibility of decisions by latent fingerprint examiners. *PLoS* 7, e32800. <https://doi.org/10.1371/journal.pone.0032800>.
- van Baar, R.B., van Beek, H.M.A., van Eijk, E.J., 2014. Digital forensics as a service: a game changer. *Digit. Invest.* 11, 54–62. <https://doi.org/10.1016/j.diin.2014.03.007>.
- van Beek, H.M.A., van Eijk, E.J., van Baar, R.B., Ugen, M., Bodde, J.N.C., Siemelink, A.J., 2015. Digital forensics as a service: game on. *Digit. Invest.* 15, 20–38. <https://doi.org/10.1016/j.diin.2015.07.004>.
- van Beek, H.M.A., van den Bos, J., Boztas, A., van Eijk, E.J., Schram, R., Ugen, M., 2020. Digital forensics as a service: stepping up the game. *Forensic Sci. Int.: Digit. Invest.* 35, 301021. <https://doi.org/10.1016/j.fsidi.2020.301021>.
- Watkins, K., McWhorte, M., Long, J., Hill, B., 2009. Teleporter: an analytically and forensically sound duplicate transfer system. *Digit. Invest.* 6, 43–47. <https://doi.org/10.1016/j.diin.2009.06.012>.
- Wilson-Kovacs, D., 2019. Effective resource management in digital forensics. *Policing: Int. J.* 43 (1), 77–90. doi:10.1108/pijpsm-07-2019-0126.

Article 3: Sunde, N. (2022). Unpacking the evidence elasticity of digital traces. (Manuscript submitted for publication).

Unpacking the evidence elasticity of digital traces

Nina Sunde

Abstract

Digital traces are sought in most criminal investigations, due to their usefulness in shedding light on the case circumstances and evidential themes. They are often trusted to be value-neutral and credible. However, a recent study on digital forensic (DF) decision-making revealed that the human factor greatly influences the construction of digital evidence. The study found that DF practitioners were biased by contextual information and produced inconsistent results during DF casework. This article applies a qualitative lens to explore how the statistically determined variance materialises in DF reports. The article examines the role of interpretative flexibility when the DF practitioner constructs the digital evidence. It develops ‘evidence elasticity’ as a concept for describing the mutability of digital traces as knowledge objects. The article explores the role of evidence elasticity for constructing narratives involving digital evidence and how this sometimes may result in misinformation with the propensity to mislead actors in the criminal justice chain, such as the investigators, prosecutors, or judges.

Keywords: Digital traces, digital evidence, digital forensics, criminal investigation, interpretative flexibility, elasticity

Introduction

Digital information from various sources such as mobile phones and laptops is frequently seized in criminal investigations. According to the Digital Forensic Strategy issued by the UK National Police Chiefs Council, over 90% of all recorded crimes have a digital element (NPCC 2020). Digital traces are transformed into evidence through the digital forensic process, which is an iterative multistage process for the identification, collection, examination, analysis, and presentation of the evidence (Flaglien 2018). The digital forensic process is primarily conducted by personnel – the digital forensic (DF) practitioners - with adequate expertise to ensure a forensically sound transformation (McKemmish 2008).

Despite the available knowledge about the technical sources of errors in digital traces and the necessity of error mitigation (e.g., Casey 2002, SWGDE 2018), digital evidence is often

trusted to be credible and value-neutral representations of the truth (e.g., Kaufmann 2017, Erlandsen 2019, Reedy 2020). According to Brookman *et al.* (2020, p. 23), ‘the “voice” of science and technology seems to carry greater weight than other forms of information (in the minds of criminal justice actors at least) in criminal justice narratives of this kind’. Such presumptions might be related to common techno-fallacies: beliefs that technology is neutral, that ‘facts’ speak for themselves, or the belief in a hundred per cent fail-safe systems (Marx and Guzik 2017). These beliefs may rest on assumptions of ‘mechanical objectivity’ (Daston 1992, Daston and Galison, 2007), which implies that machines produce richer, better, and truer evidence than humans. The trust in digital evidence is even reflected in common law, for example, in England and Wales, where the rule that governs the admissibility of electronic evidence states a presumption that computer systems are reliable: ‘in the absence of evidence to the contrary, the courts will presume that mechanical instruments were in order at the material time’ (The Law Commission 1997, para 13.13.). The problem is, however, that, if one believes that digital evidence is value-neutral, objective and credible factual representations of truth, it may also be assumed to be free from technical and human error. The consequence of such a misconception may be less scrutiny and quality control. Flawed or misleading digital evidence may have major consequences on people’s lives. It may lead to erroneous charges and arrests in the pre-trial phase, as well as wrongful convictions (e.g., Smit *et al.* 2018).

The DF discipline has largely focused its attention on technical sources of error (e.g. Casey 2002, SWGDE 2018) in order to prevent *evidence dynamics*, which are described as ‘any influence that changes, relocates, obscures, or obliterates evidence, regardless of intent between the time evidence is transferred and the time the case is resolved’ (Casey 2011, p. 27). Evidence dynamics relates to material changes in the evidence (Casey 2011), and the concept does not encompass factors influencing the digital trace as a knowledge object. As this study will show, even if the evidence integrity is safeguarded and maintained from start to end, the evidence may still be constructed differently due to the *interpretative flexibility* (Collins 1981, Doherty *et al.* 2006) of digital traces. Kruse (2021) argues that, even if the objects such as trace evidence are stable enough to be moved between epistemic cultures (Knorr-Cetina, 1999) in the criminal justice system, the knowledge they are meant to move does not necessarily remain stable, and that this *mutability* is problematic in a legal context. Ask *et al.* (2008, p. 1247) use the term *elasticity* to describe ‘the level of ambiguity associated with a piece of information, which leaves room for subjective interpretations’. Since the

article explores the mutability of digital traces resulting from the subjective interpretation by DF practitioners, it is further referred to as evidence elasticity.

Due to the subjective nature of the DF investigative process, it is of particular value to gain insight into how digital traces may be turned into *misinformation*, which, in a worst-case scenario, may mislead the legal decision-makers and result in miscarriages of justice. The article draws on S e (2019)'s unified account of information, misinformation, and disinformation, which considers the information to be *misleading* when it has the propensity to cause false beliefs (S e 2019). The article centres on *misinformation*, which entails that the information's misleading nature is unintended and the result of an honest mistake, such as an unconscious bias, as opposed to *disinformation*, which is intended to mislead the recipient (S e 2019, p. 5946). *Evidence elasticity* and *misinformation* are discussed in relation to legal concepts for assessing evidential value, as described by Anderson *et al.* (2005). The aim is to explore whether and how the evidential value is crafted by the DF practitioner and to shed light on how the evidence elasticity enables the DF practitioner to turn the traces into misinformation with the propensity to mislead the legal decision-maker's assessment of evidential value.

The article is structured as follows: First, related research and the theoretical perspectives are outlined, followed by a description of the applied methods. The analysis section presents and discusses the elasticity of digital traces and explores its range and function for constructing digital evidence and the narratives involving such evidence. Then, evidence elasticity is discussed in relation to the legal concepts for assessing evidential value to shed light on how misinformation may influence the legal decision-maker's assessments concerning the relevance, credibility, and inferential/probative force or weight of digital evidence.

Related research and theoretical perspectives

Scientific practice and knowledge construction have been the subject of several empirical studies (e.g., Latour and Woolgar 1979, Knorr-Cetina 1982), and several studies within the Science and Technology (STS) scholarly tradition have explored the construction of forensic evidence (e.g., Cole 2001, Dahl 2009, Kruse 2016). There is a sparse but growing body of empirical research on various aspects of DF practice, such as acquisition tasks (Carlton 2007, Hewling 2013), preview and triage (James and Gladyshev 2013, Wilson-Kovacs 2019, Rappert *et al.* 2021), investigative strategies for content analysis (Haraldseid 2021), reporting practices (Sunde 2021, Stoykova *et al.* 2022), quality procedures (Jahren 2020, Tully *et al.*

2020), and collaboration practices and professional dynamics between investigation officers and DF practitioners (Hansen *et al.* 2017, Wilson-Kovacs 2021).

Ask *et al.* (2008) explored the asymmetric scepticism related to different evidence types and found interpretative flexibility (elasticity) in police trainees' assessments of the credibility of digital evidence (CCTV evidence). Brookman *et al.* (2020, 2021) studied how digital evidence such as CCTV evidence contributed to identifying and charging suspects in homicide investigations and the risks associated with such evidence. They identified packaging for court as a highly constructive process, which involved decisions about which images to include and exclude for court, ordering and highlighting important information. This article advances these insights by exploring the elasticity of digital traces mediated by DF practitioners.

Former studies of criminal justice narratives have followed the evidence and the developing narratives through the justice chain (or parts of it) (e.g., Dahl 2009, Santos 2014, Kruse 2016, Lam 2016, Costa and Santos, 2019, Offit 2019, Brookman *et al.* 2020). The novelty of this research relates to exploring the multiple narratives constructed and 'framed' (Goffman 1975) based on the same evidence file.

The analysis draws on concepts from Actor-network theory (ANT), which views human and non-human actors as actors with agency (Latour 2005). The DF practitioner is at the focal point of the article. Central non-human actors in this study are the *assignments* provided to the DF practitioners and the *reports* produced as a result of the DF process, where the findings are represented in text or visualisations. In ANT, the concept of *inscription* entails turning matter into symbolic forms, such as text, graphs, diagrams, and lists (Kruse 2016). This article addresses the shortcomings of the evidence dynamics concept by drawing particular attention to the function of the evidence elasticity when the DF practitioners turn the digital traces into inscriptions in DF reports.

Material and methods

The article is based on empirical data from DF reports obtained in research, involving a quasi-experimental design (further referred to as 'the DF experiment') with 53 DF practitioners from eight countries (Sunde and Dror 2021). The primary purpose of the DF experiment was to explore contextual bias and inter-practitioner reliability in DF decision-making. In short, the DF practitioners were given a vignette with a scenario description, an assignment, and a mock evidence file (see Appendix 1 in Sunde and Dror 2021). They were asked individually

to analyse the evidence file and document their findings in a report. According to the vignette, the evidence file was secured from the laptop of Jean Jones, CFO of the firm, M57.biz. There had been an information leakage from M57.biz, which involved the posting on a competitor's web forum of confidential information about the employees. The information allegedly originated from Jean Jones' laptop. The participants were asked to find out what had happened, clarify Jean Jones' involvement in the incident and write a report on the result.

To explore whether they were prone to contextual bias, the participants received different contextual information indicating guilt and innocence in the vignette. A control group received only the basic scenario. The inter-practitioner reliability was also explored for each experiment group. For details about the design and methods applied in the DF experiment, see Sunde and Dror (2021). The results of the DF experiment indicated that contextual information led to biased observations, i.e. systematic differences in how many traces each of the DF practitioners discovered and deemed relevant to include in the report. Further, the results showed low reliability in inter-practitioner observations, interpretations and conclusions

(Sunde and Dror 2021).

This article applies a critical qualitative lens to explore how the variation and inconsistency in the results came about in the DF practitioners' reports from the DF experiment. A thematic analysis method (Braun and Clarke 2006) was applied, through a combined deductive and inductive approach in QSR International NVivo Pro Edition, Version 12.1.1.256 64 bit. Some themes were already triggered during the initial coding for the statistical analysis of the DF experiment, which guided the deductive analysis approach. The inductive analysis focused on additional variance elements not directly related to the statistical measurements. Table 1 presents the codes and themes emerging from the analysis, which also form the *Results and analysis* section's structure.

The analysis centred on the diversity of the descriptions, and the frequency of traces is only mentioned when it is considered relevant for the elasticity or quality aspects (e.g., false positives). Included quotes from the DF practitioners' reports are marked with 'D' (e.g., D56). An English translation is offered when the original quote was in the Norwegian language.

Table 1: Overview of identified codes and themes, and in which section they are presented

Themes	Codes
Interpretations of the assignment	<ul style="list-style-type: none"> - Assignment consistent with the original - Assignment differing from the original - Missing account of the assignment
Reconstructing what had happened	<ul style="list-style-type: none"> - Search for and/or observations of the leaked spreadsheet (m57plan.xls) on the suspect's computer - Observing the file m57biz.xls on the suspect's computer - Comparing the leaked file (m57plan.xls) with the file m57biz.xls found on the suspect's computer
Determining who were involved	<ul style="list-style-type: none"> - Activity descriptions related to entities - Activity descriptions related to persons
Constructing conclusions	<ul style="list-style-type: none"> - What? <ul style="list-style-type: none"> o CFO fraud o Phishing/spear phishing o Hacking o There was no information leakage - Who? <ul style="list-style-type: none"> o Jean had caused/contributed to the information leakage o An insider caused the information leakage o An outsider performed the information leakage - How? <ul style="list-style-type: none"> o The info was sent by email o The info was copied out on a USB o The info was obtained through hacking o No information leakage had happened

Results and discussion

The DF process begins with identifying and collecting devices with data, typically smartphones or laptops. Due to established principles for forensically sound handling, the device is not accessed immediately. Instead, a DF practitioner acquires a copy of the device's content (referred to as an evidence file) and performs the subsequent analysis on the evidence file. Software is used to process the evidence file, enabling the DF practitioner to explore the content as meaningful information. After conducting the analysis, the DF practitioner writes a report on how the DF investigation was performed and their findings.

This study follows the digital traces through the analysis and presentation stages of the DF process (Flaglien 2018). It starts at the point when the DF practitioner receives the assignment and the copy of Jean Jones' laptop (the evidence file).

Key objects and names from the DF experiment are presented in Table 2.

Table 2: Overview of central names and objects referred to in the analysis

Names and objects	Explanation
The evidence file	The accurate copy of the information on the suspect's computer. File name: nps-2008-jean.E01
Jean Jones	The CFO of the company M57.biz, user of the seized computer, referred to as 'the suspect's computer' or the 'evidence file'
Alison Smith	The manager of the company M57.biz
M57plan.xls	The allegedly leaked spreadsheet
M57biz.xls	Spreadsheet found on the suspect's computer at two locations: a) the Desktop folder b) the email container file (Outlook.pst)

First, the DF practitioner's reproduction of the assignment is explored, followed by an examination of their descriptions of central digital traces and events found on the evidence file. Finally, their reported conclusions are analysed and discussed.

Interpretations of the assignment

The DF practitioner typically receives an assignment from the tactical investigation team, which guides the scope of the DF analysis (Sunde and Dror 2021). In the DF experiment, all participants received the same assignment in writing: 'What has happened, and what was Jean's involvement in the reported incident?' The assignment is also the article's analytical starting point. Despite receiving similar assignments in writing, a correct account of the assignment was found in approximately half of the DF reports, and almost one-fifth of the reported assignments diverged substantially from the original assignment (Sunde 2021, p. 591). Examples of the varying assignments stated in reports are included in Table 3.

Table 3: Original assignments versus examples of reported assignments

Original assignment	Examples of reported assignments
What has happened, and what was Jean's involvement in the reported incident?	<ul style="list-style-type: none"> - 'to find out how internal business information was given to a competing business corporation' (D2). - 'to determine whether there were files/material on the evidence file that could provide information about the suspect's role in this matter' (D6). - 'whether there was a data breach at the firm M57.biz and Jean Jones involvement in the data breach' (D31). - 'to perform a content analysis of the evidence file, and to document any findings that may shed light on the case' (D48).

Kruse (2016) points out that, although science and crime scene investigations have much in common, there is a fundamental difference in their purposes: Research aims to contribute to the growing body of generalised knowledge, while forensic science aims to produce knowledge about the particular. The assignment defines what *in particular* the DF practitioner should produce knowledge about, and the deviant accounts of the assignment demonstrate that a written assignment does not guarantee that the sender and the receiver read the same meaning into it. The deviating accounts of assignments are an important finding. They demonstrate that it is not the assignment itself but the DF practitioners' *translation* of it that guides their further discovery process and assessments of what traces are relevant or irrelevant to the investigation. This shows that the DF practitioner may start the analysis with a different assignment in mind than the one they originally received, highlighting the importance of DF practitioners documenting the assignment in DF reports, to enable scrutiny and detection of misinterpretations that may have led the DF investigation astray.

Reconstructing what had happened

Digital evidence is co-constructed within what may be referred to as a *socio-technical expert system*, where highly skilled DF practitioners perform physical and cognitive tasks in interaction with powerful hardware and custom-designed software. The technology extends their cognitive abilities and helps them make sense of digital information collected during an investigation (see Brey 2017). They operate within a cultural environment, with domain-specific processes, procedures, and defined goals. The DF expertise, combined with access to the necessary technology, makes the DF practitioner an obligatory passage point (OPP) (Callon 1986) for digital evidence in a criminal investigation.

The DF practitioner plays an active role in the DF process and makes numerous judgements and decisions that impact which findings are relevant to report, how to assess, control, and document their credibility, and how to evaluate their evidential value (Sunde 2021, 2022, Sunde and Dror 2019, 2021). In their reports, the DF practitioners present *findings*, which are an overview of what they have deemed information relevant to the investigation. They also provide *opinions*, which are statements about what the traces are, what they mean and their evidential value. The original data (the input) and the reported results are observable by others. However, the cognitive processes involved in the judgements and decisions are invisible and may thus be perceived as a black box. The counterfactual experimental setting provided a unique opportunity to observe the consequences from diverse interpretations of a

similar evidence file, through the study of the reports produced by the 53 DF practitioners participating in the DF experiment.

The DF experiment vignette indicated that the allegedly leaked spreadsheet with confidential information was of particular interest. The DF practitioners were informed of the file name 'm57plan.xls', and that the file had been discovered on the competitor's web forum. A snapshot of the leaked spreadsheet was also enclosed. To determine whether the file was leaked from the suspect's computer, an essential investigative step would be to establish whether there were traces of the leaked spreadsheets on the suspect's computer. The DF practitioners could deploy several strategies to find the file, for example, a text string search for the filename or text/content in the spreadsheet or a search for all spreadsheets in case the spreadsheet had been renamed. The results related to this issue are discussed below.

[The leaked spreadsheet \(m57plan.xls\) was/was not on the suspect's computer](#)

Since a mock evidence file was used, the ground truth was known, although not to the DF practitioners participating in the DF experiment. The leaked spreadsheet was, in fact, *not* on the suspect's computer, and a correct approach would therefore be to document the search strategy and result: that m57plan.xls (the leaked spreadsheet) was not found on the computer. Approximately a quarter of the DF practitioners provided a report corresponding with the ground truth and stated explicitly that a spreadsheet named M57plan.xls was not found on the suspect's computer (Sunde and Dror 2021). However, even though the leaked file was not on the suspect's computer, some DF practitioners (n=3) stated that it *was present* and related it to the information leakage, for example: 'The spreadsheet m57plan.xls was identified as the allegedly leaked spreadsheet. The evidence file did not contain any other relevant spreadsheets' (D29). Such a statement may be characterised as misinformation with the potential to mislead the legal decision-maker.

Those who commented that they did not find the leaked spreadsheet stated this in two ways, which illustrates the DF practitioners' different interpretations of similar information. Some stated that the spreadsheet was searched for, but *not found*, such as 'In the task, a file named "m57plan.xls" was mentioned. This file was not found on the evidence file' (D9) or 'A text string search was performed on "m57plan.xls". This search gave 0 hits' (D32). This description did not exclude that the leaked spreadsheet could be there and is in line with the well-known principle that *absence of evidence does not entail evidence of absence*.

Nevertheless, others *excluded the presence* of the leaked spreadsheet on the computer in their description, for example: ‘The file as mentioned in the report, i.e. m57plan.xls, was not available in the image file nps-2008-jean.E01’ (D53). Based on the scope of the analysis performed in the DF experiment, this is an overstatement beyond justification, which would fall into the category of misinformation.

The examination and analysis of the information obtained from a digital device may have many similarities with the crime scene investigation of a physical crime scene. Kruse (2016) found that the main contribution to the pre-trial investigation from crime scene investigators was written crime scene reports. Kruse describes: ‘When a report replaces a crime scene, the crime scene becomes more understandable through the interpretation the crime scene technicians have made, but it also becomes more limited. The details that have not been documented are lost with the crime scene, and the report shapes how the readers understand the crime scene and thus the forensic evidence’ (Kruse 2016, pp. 95-96). This description is transferable to the digital crime scene and the analysis report’s function. The DF practitioners turn the digital crime scenes into inscriptions, by describing or visualising the relevant digital traces into reports and documenting their credibility. A mediated picture of the digital crime scene is painted in DF reports based on the traces considered relevant to the case.

However, compared to a physical crime scene, there is a fundamental difference. Under optimal conditions, an accurate copy of the digital crime scene can be secured, stored, and re-examined later, which is a great advantage from a quality perspective. Invalid results, such as false positives and overstated opinions, could be corrected during quality measures such as a verification review or re-examination (Horsman and Sunde 2020). Unfortunately, research has shown that quality measures such as peer review are rarely undertaken as part of the DF process (Page *et al.* 2019, Jahren 2020). The misinformation should thus be of concern, since, in *actual casework*, it would lead to a stronger (but invalid) underpinning of the suspicion against the suspect and potentially result in an unfair administration of justice.

[Comparing the leaked spreadsheet \(m57plan.xls\) with the file on the suspect’s computer \(m57biz.xls\)](#)

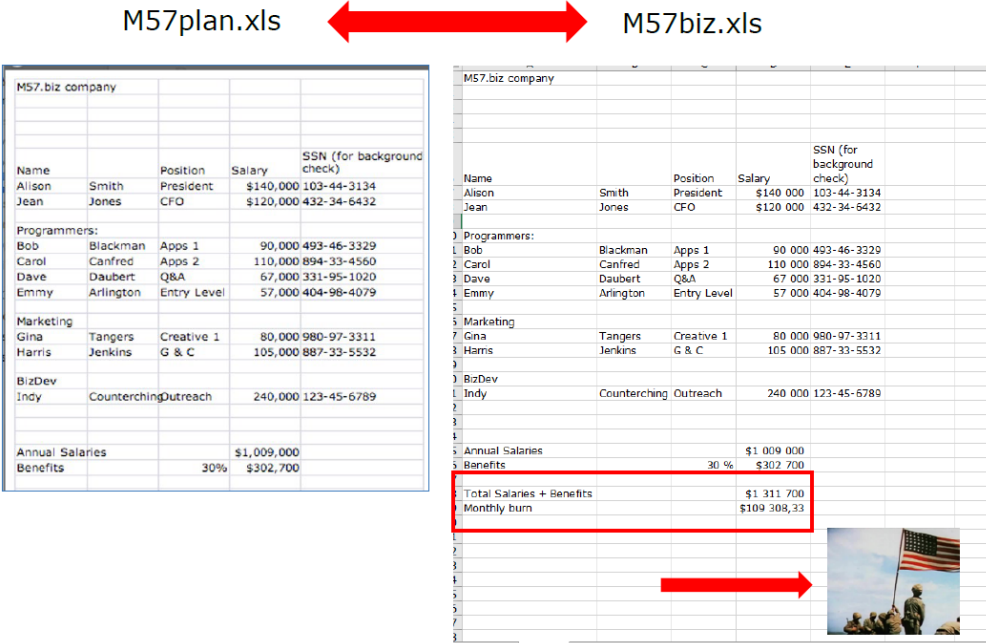
As mentioned, the leaked spreadsheet was not on the suspect’s computer. There were, however, two identical spreadsheets named m57biz.xls on the suspect’s computer, which had many similarities with the leaked spreadsheet. One was stored in the ‘Desktop’ folder and the other in the email container file (outlook.pst). They contained much of the same information

as that of the leaked file (m57plan.xls), but there were three distinct differences (see Figure 1):

- The filenames were different
- The file, m57biz.xls, had two extra rows of information, compared to the leaked spreadsheet
- The file, m57biz.xls, contained a picture of soldiers holding a flag, while this picture was not visually present in the leaked spreadsheet

According to the statistical analysis, almost all (94%) of the DF practitioners found at least one of these (Sunde and Dror 2021). An appropriate next step would be to compare the file(s) with the leaked file. Corresponding files would support a hypothesis stating that the seized computer was the source of the leaked information. As shown below, the reports included a wide range of descriptions, which illustrate the interpretation of the findings relevant to this evidential issue.

Figure 1: The leaked spreadsheet (to the left) juxtaposed to the spreadsheet located on the suspects' computer (to the right). Differences are indicated in red.



The spreadsheets were different

After comparing the files, some explicated (correctly) that the files were *different* and described the dissimilarities, for example: ‘The file of interest, with another filename than provided in the incident report, was found on the unit: M57.biz.xls’ (D57) or:

Then I examined all the xls files on the evidence file and found an xls file named ‘m57biz.xls’. This file was copied and opened in excel. The file was similar to the one described as ‘m57plan.xls’, except that it contains 2 extra rows at the end of the document and an image. The image looks like the iconic image of ‘Rising the Flag’ at Iwo Jima during the Second World War. (D6)

The spreadsheets were similar

Even *true* information may sometimes also become misinformation: ‘Through unintended omittance it is possible to have instances of true misinformation—i.e., what is said is true, but by unintentionally omitting other crucial parts of the picture it becomes misleading’ (Søe 2019, p. 5939). Common to many of the descriptions was the focus on *similarities and correspondence with the leaked file*, while downplaying or excluding the differences, for example, ‘An Excel document was discovered, which largely corresponds with the document referred to in the incident report, ref 5.12 Documents’ (D1).

Others focused on how the file *content* corresponded, such as ‘The same information as the information that had gone astray was found on Jean’s computer with the filename “m57biz.xls”’ (D18) or ‘The spreadsheet “m57biz.xls” with corresponding content as what was posted was found on the computer in the user folder of user “Jean”’ (D37). Others described whether the spreadsheets *visually* looked similar:

A document named ‘m57biz.xls’ has the same content as the document ‘m57plan.xls’. [...] The document is stored on Jean’s computer with file path ‘\Documents and Settings\Jean\Desktop\m57biz.xls’ and has the md5 hash of ‘e23a4eb7f2562f53e88c9dca8b26a153’. The document corresponds visually with the document shared on the internet. (D38)

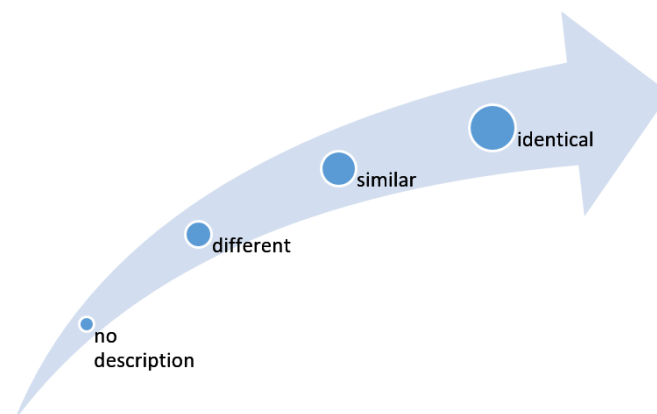
Since the descriptions present only some of the relevant information, they are inaccurate and incomplete and can potentially mislead the legal decision-maker.

The spreadsheets were identical

Some stretched their accounts of the similarities even further and described the content as identical, which is an invalid interpretation, as shown in Figure 1. For example, ‘I did not find the file M57plan.xls on the evidence file, but I found M57biz.xls with identical content’ (D5) or ‘20.07.2008 01:28:00, Jean responds to the email and encloses the file m57.biz.xls. Through the search for the exact filename on the evidence file, the original file was discovered, which is identical to the file that was leaked on the internet’ (D31).

The variation in descriptions of the two spreadsheets, from *different* to *identical* (see Figure 2), demonstrates the *range of elasticity* of the digital traces. Kruse underlines that constructed does not entail that they are not real, but rather that they could be constructed differently (Kruse 2016, p. 7), and the 53 reports clearly demonstrate how the elasticity becomes a construction tool for crafting and customising coherent building blocks for an emerging narrative. By describing the spreadsheets as similar or identical, without commenting on the differences, the DF reports can mislead legal decision-makers into believing that the leaked document was found on the suspect’s computer or that a causal relationship between the two documents was established. Such descriptions may fuel an illusion of incriminating evidence against the suspect.

Figure 2: The range of descriptions of the relationship between m57plan.xls and m57biz.xls.



Determining who were involved

Digital traces obtained from computers and mobile phones are vital for reconstructing events and activities. However, they have little value if not linked to a person. Since multiple users may access a computer, attributing these activities to a person, with the required level of certainty (source level issues), remains a recurring challenge in DF investigations (see, e.g., Cohen 2013). In other words, proving who had their fingers on the keyboard and performed certain activities that left traces on the digital device can be complicated. Another complicating factor is that, in addition to excluding other possible suspects as sources of the traces, the DF practitioner must determine whether human activity or a machine/system generated the traces. For example, suppose illegal imagery is found on a computer. In that case, the DF practitioner must justify that it was downloaded by a particular person and refute that it was not the result of an automated procedure happening in the background when the user visited a webpage.

Similar to a typical DF investigation, the participants in the DF experiment received limited and incomplete case information before starting the analysis. Although they were informed that the evidence file was secured from Jean's computer, it could not be stated with certainty who (person) had conducted the activities on the computer. There were several active user accounts on the computer, with traces of activity around the time of the alleged information leakage and traces indicating that the computer could have been compromised through, for example, hacking.

Due to the insufficient information and the uncertainty concerning who caused the traces, the DF practitioners ought to describe *entities* (e.g., user accounts, email addresses) rather than *persons* related to the alleged activities or events (e.g., sending email, searching the internet). For example: 'According to the internal metadata, the document was created on 12.06.2008 17:13:51 by a user named Alison Smith and last saved on 20.07.2008 at 03:28:03' (D1) or email addresses 'An email sent to jean@m57.biz on 21.07.2008 at 01:43:19 from bob@m57.biz discusses the publishing of the attachment which is reported' (D49). Such an approach does not implicitly assume who performed the activity. In contrast, many described activities performed by a *particular person*, for example stating that Jean replied to emails: 'Jean Jones replies to the last inquiry and encloses the document m57.biz.xls (item # 46654)' (D15) or that she created the document: 'Jean creates this document and sends it to the malicious email address. She is not aware that the recipient is not Alison, and it seems like she

has no intentions of sharing this information on purpose' (D47). Another described that Jean sent the spreadsheet and altered the document:

It is proven that Jean Jones sent the file to an email address outside their own firm domain. [...] Examination of the excel file shows that Jean Jones was the last person who edited the file. [...] Jean Jones receives an email from (the real Alison Smith – based on the email address of the firm domain) about strange things happening. (D2)

The analysis shows that the DF practitioners' mediations relate not only to the digital traces but also to who was involved. Here, the elasticity relates to how far the DF practitioner goes in the trace-linking process, that is whether they link the trace to the entity or make an additional inferential leap, and link from trace to person. When the reasoning and justification of the relationship between trace and entity/person lacks transparency, the legal decision-maker may be misled into believing that the link between a suspect and a criminal activity is more certain than it really is.

Constructing conclusions

When zooming in on a digital trace, it is often an assembly of several components. The email sent from Jean's computer with the enclosed spreadsheet, m57biz.xls, consists of content and metadata. However, when zooming in on the metadata, another assembly of temporal and spatial traces of transportation of the information from sender to receiver via several servers emerges. The observed traces are rarely viewed in isolation; they are ordered, linked, and related to the context of the case under investigation. When traces are pieced together during the abductive reasoning process, they may enable several plausible explanations (Anderson *et al.* 2005). These explanations form sub-narratives, which may be further assembled into narratives of what has happened, how it was done, who was involved, why they did it, when and where. Anderson *et al.* (2005) describe this process as 'connecting the dots'. The number of traces and factors such as ambiguity and completeness influence the forming of explanations and narratives (Tecuci *et al.* 2016, pp. 159, 163).

Several plausible scenarios could be developed based on the contextual information and the digital traces on the suspect's computer; for example, the confidential information could have been leaked by

- copying it onto a thumb drive
- uploading it to the internet

- breaking into the system and stealing it
- tricking Jean into sending it by email (phishing)

There were several possible perpetrators, such as Jean, another insider, an outsider, or an insider and an outsider in collaboration. The evidence file from Jean Jones' laptop contained traces consistent with all these scenarios. Yet, none could be refuted on the basis of the information on the evidence file obtained from the suspect's computer. The different observations and interpretations in the M57 case led to many different conclusions or summaries of the results. Some DF practitioners concluded that Jean had *caused or contributed* to the information leakage. For example, 'If the user account "Jean" on the analysed system is Jean Jones, the findings mentioned here indicate that Jean may have sent the file "m57biz.xls" per email to tuckergorge@gmail.com or alison@m57.biz on 2008-07-20 03:28:00 CEST' (D42) or 'The employees involved are Alison SMITH, manager of M57.biz, and Jean JONES, CFO of M57.biz. Additionally, email communication would also suggest that tuckergorge@gmail.com also has an involvement in the leakage of confidential information' (D62). Others concluded, on the contrary, that the spreadsheet was *not* sent/shared from the computer, for example, 'Based on the analysis referred to under section 5, no information was found indicating that the spreadsheet, m57biz.xls, or other documents with the same content were shared from the unit' (D28).

Many indicated that the leaked spreadsheet could be associated with Jean sending the spreadsheet outside the firm, but they highlighted that she was *deceived*, which indicates innocence, for example, 'It seems like the firm m57.biz was exposed to CFO-fraud' (D19) or 'The examination has uncovered that Jean Jones appears to have been the victim of a spear-phishing attack, where she has sent the leaked document to an attacker, who has pretended to be her colleague, Alison Smith' (D15).

Some provided *alternative explanations* of the information leakage, such as the mounting of a USB and the installation of VMWare, for example, 'There are traces of a thumb drive being connected to the machine just before the above-mentioned email was sent. Further examinations should be undertaken to determine whether the spreadsheet might also have been shared from this thumb drive' (D22) or:

The software VMWare was found, which makes it possible to run a virtual machine on the computer. The programming tool C++ was also discovered. Both were installed just

before the email was sent, and both programs were uninstalled. It is therefore not possible to eliminate that Jean's computer has been compromised in some way. (D34)

While most of the conclusions associated the uncovered traces with Jean being tricked in a phishing attack, others proposed the *insider theory* to explain the information leakage, suggesting one of the programmers or the manager as possible perpetrators. For example, 'It should also be mentioned that it was Alison Smith who reported the incident. It should be investigated with the same strength whether Alison Smith may have an agenda to blame Jean Jones' (D2).

The analysis of conclusions shows that sense-making and linking of traces may lead to many different explanations of what happened and who was involved, which may be referred to as a 'story', 'narrative' or 'scenario'. The story model of Pennington and Hastie (1993) offers descriptions of how people reason when they are exposed to a lot of information or evidence and how they create, evaluate, and select scenarios. A scenario is anchored in common sense generalisations (Crombag *et al.* 1994) and consists of at least a central action and a scene that makes the central action understandable (Van Koppen and Mackor 2020). Scenarios may also have elements such as a scene, motive, actor, and consequences (Pennington and Hastie 1993). The DF practitioner contributes to the narrative of the crime under investigation by linking the digital traces and presenting narratives or narrative sequences inferred from the evidence and general world knowledge. The narrative is essential for making sense of information and deciding whether it is relevant as evidence (Stubbins and Stubbins 2008).

One cannot simply attribute criminality to individual events, and one cannot easily attest to individual persons as offenders. If someone is to be identified as an offender, it is above all (and beyond purely legal arguments) necessary to have a story at hand that convincingly and compellingly justifies this attribution. (Dollinger and Heppchen 2019, p. 255)

At the same time, narratives – even those with dubious validity - may appear attractive and appealing (Stubbins and Stubbins 2008). Since many of the conclusions in the DF experiment are articulated as narratives or sub-narratives, they may have the potential to lead (or mislead) an investigation in many different directions.

Although Goffman's (1975) framing theory involves face-to-face interactions, it is helpful to understand the power associated with narrative construction. The DF practitioners influence the report reader's attention, by highlighting which traces or issues are relevant (Goffman's in-frame) and downplaying or omitting what is irrelevant (Goffman's out-of-frame). The DF practitioners order, describe, and mediate the narrative surrounding the traces. Present in the report are not the original traces but material-semiotic objects of knowledge, which are an 'active, meaning-generating axis of the apparatus of bodily production' (Haraway 1991, p. 200). The text thus becomes an active and transformational entity. As Asdal (2015) describes, turning something into an issue might also entail becoming a non-issue, a question to be handled by certain issue experts. In the context of DF casework, the expert describing the issue may be one of the few with technical expertise to fully understand the traces and their context. Since other actors in the criminal justice chain may lack the expertise to challenge the findings in a meaningful way, even flawed descriptions of traces and invalid opinions concerning these may become a non-controversy or a non-issue and be assigned factual status.

The variable conclusions from the DF experiment may, on one hand, be viewed as a challenge to investigation quality. Since some of the trace interpretations and conclusions observed in the current study were erroneous or one-sided, they could, in an actual investigation, have a 'misframing' (Goffman 1975, p. 308) potential, by introducing contextual bias on subsequent investigative tasks, and lead the investigation astray. Under such conditions, the trace and conclusion elasticity would challenge the fair administration of justice and the rule of law. On the other hand, the various findings and conclusions may be perceived as a means to broaden the scope of the investigation by illuminating the multiple explanations of the traces and the potential lines of inquiry. However, in an actual criminal investigation, only one DF practitioner, and occasionally a few DF practitioners, would perform the analysis and write a report (e.g., Jahren 2020). The elasticity would thus not be visible.

Mol (2002) suggests that there are multiple versions of the same body in medicine: a 'body multiple', where the same diagnoses were enacted differently by various medical disciplines. The DF reports provide insight into the DF practitioner's construction process of a new object – the digital evidence – which is an assemblage of material-semiotic representations of the traces, disseminating what they are, what they mean, and what story they tell concerning the case under investigation. The traces in themselves are meaningless, but, ordered, placed in a context, explained, and evaluated, they become meaningful pieces of information (Edmond 1999). The transparency of the DF practitioners' enactment and mediation depends on how

accurately they document the applied procedures and are often black-boxed when the report is finalised. For digital evidence, plurality not only relates to various disciplines enacting the same ‘diagnosis’ differently but also includes representatives of the same discipline enacting, translating, and inscribing the digital evidence differently into DF reports, by using elasticity as a construction tool, which enables a *digital evidence multiple*.

The empirical data collected from the DF experiment sheds light on aspects that are rarely visible in actual casework. It provides a unique opportunity to identify which components of the digital evidence are elastic, the range of the elasticity, and how the elasticity functions as a construction tool to develop plausible and coherent explanations and narratives. The article demonstrates how evidence elasticity enables the construction of multiple versions of the same traces, interpretations of traces, and various conclusions based on these. The DF discipline has centred attention on preventing *evidence dynamics* (Casey 2011) but has lacked a concept encompassing possible misleading aspects related to the digital trace as a knowledge object. The evidence elasticity concept complements the concept of evidence dynamics. In context, the evidence dynamics and evidence elasticity can explain the misinformation and flaws that may relate to the physical/digital aspects of the evidence and those related to the interpretation of the evidence as a knowledge object.

Evidence elasticity and the legal criteria for evidential value

A piece of evidence has no value in isolation – it must be considered in light of a questioned matter and conditioning information. The DF practitioners’ position as experts provides them with definition power, and, when reporting the findings, they also mediate how the traces should be understood in light of the case under investigation. Discussing these aspects in relation to criteria for evidential value is helpful for understanding whether and how the mediation also encompasses the evidential value, and whether the elasticity may be problematic in such a context. The discussion centres on the assessment of evidential value, performed for example during the investigation or in court, and does not address the issue of admissibility, the legal threshold for admitting forensic evidence to the trial. Anderson *et al.* (2005) relate the evidential value of tangible (as opposed to testimonial) evidence to three components: reliability, credibility and probative/inferential force or weight (Anderson *et al.* 2005, Tecuci *et al.* 2016).

Relevance makes the evidential theme (matter to be proved) more or less probable (Anderson *et al.* 2005). The relevance of a trace often depends on other obtained traces (Tecuci *et al.* 2016). In the M57 case, the spreadsheet on the suspect's computer was assessed in light of the allegedly leaked spreadsheet found on the competitor's website. The DF practitioners mediated the relevance by establishing a strong or weak link between findings on the suspect's computer and the allegedly leaked spreadsheet, or between inculpatory information and the suspect.

Credibility refers to three components: authenticity, accuracy/sensitivity and reliability (Anderson *et al.* 2005). *Authenticity*, which relates to whether the evidence is a genuine representation of what it appears to be (Anderson *et al.* 2005, Tecuci *et al.* 2016) is closely related to *accuracy* or *sensitivity*, which is about whether the evidence provides sufficient resolution to discriminate between possible events/explanations (Anderson *et al.* 2005, Tecuci *et al.* 2016). The assessment relates to having enough information to infer or evaluate safely. For example, Brookman and Jones (2021) found that poor-quality CCTV footage could cause misinterpretation when observing events or attempting to identify a suspect in a homicide investigation, and was used to explain away contradictory evidence, such as the discrepancy between a suspect's coat and the colour seen on the CCTV footage. In the DF experiment, the spreadsheet from the competitor's website had been secured through the 'cut and paste' technique, which entailed the file itself and associated metadata (data about data) being unavailable to the DF practitioners. Hence, they lacked crucial and accurate information to draw inferences about its authenticity or many other investigative questions, such as who created the file or when was it created, and to compute a digital fingerprint of the file. In such situations, there is a risk that the DF practitioner unintentionally and unconsciously fills in the gaps with non-evidenced assumptions. For example, the DF practitioners consistently inferred that the spreadsheet had appeared on the competitor's forum after a spreadsheet was sent from the suspect's email; yet, there was no information underpinning the order of events.

Reliability relates to whether the evidence was produced in a repeatable, dependable, and consistent manner (Anderson *et al.* 2005, Tecuci *et al.* 2016). While Anderson *et al.* (2005) relate the reliability to the operating characteristics of the device used to generate the trace, multiple devices must be considered when establishing reliability in a DF context. The reliability assessment must encompass the device involved in generating the trace and the devices and technology used for collecting and examining the data during the DF process. However, the research shows that the reliability not only hinges on the technical instruments

generating or analysing the evidence. The results show that the ‘human instrument’ has low reliability and can produce varying descriptions of the particular traces, suggest different meanings when the traces are considered with other evidence and construct substantially divergent conclusions.

Inferential/probative force or weight of evidence is about how strongly the evidence favours or disfavors particular hypotheses or propositions (Tecuci *et al.* 2016). This issue relates largely to how the conclusions were articulated, and the analysis showed that narratives played an essential role. During narrative construction, the traces are put in a context that gives them meaning and value for the case under investigation. All conclusions in the DF reports were crafted as a single one-sided narrative, based on their findings (see also Sunde 2021). They presented various accounts of whether and how an information leakage had happened. While some constructed the narrative with the suspect as the perpetrator, others presented her as a victim of fraud. The analysis shows that, although divergent, none of the narratives stood out as implausible. Granhag and Ask (2021) outline several essential factors for a narrative to be convincing. The narrative must be plausible and able to explain all or most of the evidence. It must be cohesive and not contradict any evidence. The narrative must be complete, cover all essential parts of the crime and be arranged in chronological order. Moreover, it must be unique, and there should be no alternative narrative with an equal or higher capacity to convince. The evidence elasticity observed in the DF experiment conclusions seems to facilitate necessary adjustments that enable the traces to fit into various plausible explanations and narratives. Costa and Santos (2019) add the presence of forensic evidence to this list of convincing factors. They found that forensic evidence made the narrative appear stronger, even if the evidence was weak (Costa and Santos 2019). Due to the perception of digital evidence as facts, it may create an illusion of unquestionable anchoring points.

To summarise, the discussion shows that DF practitioners are in a unique position to mediate not only what the digital traces are and their meaning – but also their *evidential value* in a legal context.

Concluding remarks

The study explored the implications of biased observations and low reliability in DF reporting practices, by following the translation and transformation of the digital evidence through one of several OPPs in the justice system, the DF practitioner. The article explored the

interpretative flexibility of digital traces, and developed the concept *evidence elasticity* for the mutability of digital evidence as a knowledge object. Evidence elasticity adds to the concept of evidence dynamics, which relates to the physical and digital aspects that may reduce the value of the evidence. In context, these concepts provide a more complete oversight of the possible flaws and misleading aspects of digital evidence.

The most essential knowledge to take away from the study of DF reports is, first, that the digital evidence transformation starts as soon as the assignment is received. It is the *interpretation* of the assignment that guides the DF practitioner's decision-making about what traces are relevant to the case under investigation, and the selected traces condition the subsequent forming of narratives. Second, the same traces may be understood differently and assigned different evidential values. When inscribed into reports, they may sometimes be turned into misinformation, with the potential to mislead the legal decision-makers. Third, the analysis has shown that the narrative-forming process is a way of 'connecting the dots' about what has happened, how, and by whom, and that elasticity enables the same trace to contribute to very different narratives – from one extreme to another: from a narrative where the suspect is innocent to a narrative where the suspect is guilty. The analysis also shows how the DF practitioners make inferential leaps (e.g., about who did what) to create coherent and complete narratives in their conclusions.

In everyday life, framing and 'stretching' of information may be necessary for social interaction. In the context of a criminal investigation, these mechanisms may shape what the information means and its value in a legal context: their relevance, credibility and probative/inferential force or weight. As shown in this study, the power to frame also entails the power to misframe. Assigning multiple DF practitioners per case, as done in the DF experiment, is not typical. Relying on a single DF practitioner's results may, on one hand, constitute a lost opportunity to uncover the full scope of relevant traces and plausible narratives. On the other hand, such an approach may introduce the risk of one-sided or misleading conclusions and erroneous results remaining undetected and uncorrected, which may lead to wrongful convictions of innocent people, in a worst-case scenario.

The counterfactual situation of the DF experiment has provided a unique insight into evidence elasticity that would otherwise be invisible. Opening the black box and gaining insight into how digital evidence is constructed empowers other actors involved in the criminal justice

chain, such as prosecutors, defence lawyers, judges and jurors. These insights enhance their ability to scrutinise the DF process and any misinformation resulting from it, such as invalid inferential leaps and misleading opinions regarding the evidential weight of digital evidence conveyed in DF reports. Knowledge about the evidence elasticity associated with digital traces is vital for countering the fallacies about technology and digital evidence and designing appropriate quality control procedures, which may prevent misleading information from cascading further into the investigation process. Knowledge is power, and knowledge about the elasticity of digital evidence may thus be an essential means for safeguarding the rule of law.

Acknowledgements: I would like to thank Professor Helene O. I. Gundhus, Professor Johanne Y. Dahl and Associate Professor Fergus Toolan for comments and constructive criticism of former versions of the paper.

Declaration of interests: the author reports there are no competing interests to declare.

References

- Anderson, T., Schum, D., and Twining, W., 2005. *Analysis of evidence*. New York: Cambridge University Press.
- Asdal, K., 2015. What is the issue? The transformative capacity of documents. *Distinction: Scandinavian journal of social theory*, 16(1), 74–90. DOI: 10.1080/1600910X.2015.1022194.
- Ask, K., Rebelius, A., and Granhag, P.A., 2008. The ‘elasticity’ of criminal evidence: A moderator of investigator bias. *Applied cognitive psychology: The official journal of the Society for Applied Research in Memory and Cognition*, 22(9), 1245–1259. DOI: 10.1002/acp.1432.
- Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77–101. DOI: 10.1191/1478088706qp063oa.
- Brey, P., 2017. Theorizing technology and its role in crime and law enforcement. In: M.R. McGuire and T.J. Holt, eds. *The Routledge handbook of technology, crime and justice*. London: Routledge, 17–34.
- Brookman, F., et al., 2020. Crafting credible homicide narratives: Forensic technoscience in contemporary criminal investigations. *Deviant behavior*, 43(3), 340-366. DOI: 10.1080/01639625.2020.1837692.
- Brookman, F. and Jones, H., 2021. Capturing killers: The construction of CCTV evidence during homicide investigations. *Policing and society*, 32(2), 125-144. DOI: 10.1080/10439463.2021.1879075.

Callon, M., 1986. Some elements of a sociology of translation: Domestication of scallops and the fishermen of St Brieuc Bay. In: J. Law, ed. *Power, action and belief: A new sociology of knowledge?* London: Routledge and Kegan Paul, 196–223.

Carlton, G.H., 2007. A grounded theory approach to identifying and measuring forensic data acquisition tasks. *Journal of digital forensics, security and law*, 2(1), 35–56. DOI: 0.15394/jdfsl.2007.1015.

Casey, E., 2002. Error, uncertainty, and loss in digital evidence. *International journal of digital evidence*, 1(2), 1–45.

Casey, E., 2011. *Digital evidence and computer crime: Forensic science, computers, and the internet*. Amsterdam: Academic Press.

Cohen, F., 2013. *Digital forensic evidence examination*. 5th ed. Fred Cohen & Associates.

Cole, S.A., 2001. *Suspect identities: A history of criminal identification and fingerprinting*. Cambridge: Harvard University Press.

Collins, H.M., 1981. Stages in the empirical programme of relativism. *Social studies of science*, 11(1), 3–10. DOI: 10.1177/030631278101100101.

Costa, S. and Santos, F., 2019. The social life of forensic evidence and the epistemic sub-cultures in an inquisitorial justice system: Analysis of Saltão case. *Science & justice*, 59(5), 471–479. DOI: 10.1016/j.scijus.2019.06.003.

Crombag, H.F.M., Wagenaar, W.A., and Van Koppen, P.J., 1994. *Dubieuze zaken: de psychologie van strafrechtelijk bewijs*. 2nd ed. Amsterdam: Contact.

Dahl, J.Y., 2009. *DNA-det sikreste av detsikre eller...?: En sosiologisk studie av usikkerheter knyttet til bruk av DNA i strafferettspleien*. Thesis (PhD). Norwegian University of Science and Technology, Norway. Available from: <http://hdl.handle.net/11250/268114>.

Daston, L., 1992. Objectivity and the escape from perspective. *Social studies of science*, 22(4), 597–618. DOI: 10.1177/030631292022004002.

Daston, L. and Galison, P., 2007. *Objectivity*. Zone Books. Google Play Store.

Doherty, N.F., Coombs, C.R., and Loan-Clarke, J., 2006. A re-conceptualization of the interpretive flexibility of information technologies: redressing the balance between the social and the technical. *European journal of information systems*, 15(6), 569–582, DOI: 10.1057/palgrave.ejis.3000653.

Dollinger, B. and Heppchen, S., 2019. Narratives of conviction and the re-storying of ‘offenders’. In: J. Fleetwood, L. Presser, S. Sandberg, and T. Ugelvik, eds. *The Emerald handbook of narrative criminology*. Bingley: Emerald Group Publishing, 303–320.

Edmond, G., 1999. Science, law and narrative: Helping the facts to speak for themselves. *Southern Illinois University law journal*, 23(3), 555–584.

Erlandsen, T.E., 2019. *Fallacies when evaluating digital evidence among prosecutors in the Norwegian police service*. Thesis (MSc). Norwegian University of Science and Technology, Norway. Available from: <http://hdl.handle.net/11250/2617771>.

Flaglien, A.O., 2018. The digital forensic process. *In: A. Årnes, ed. Digital forensics*. Hoboken: Wiley, 13–49.

Goffman, E., 1975. *Frame analysis: An essay on the organization of experience*. Harmondsworth: Penguin Books.

Granhag, P. and Ask, K., 2021. Psykologiska perspektiv på bevisvärdering. *In: P. Granhag, L. A. Strömwall, K. Ask, and S. Landström, eds. Handbok i rättspsykologi*. Stockholm: Liber, 427–445.

Hansen, H.A., Andersen, S., Axelsson, S., and Hopland, S., 2017. Case study: A new method for investigating crimes against children. *Annual ADFSL Conference on Digital Forensics, Security and Law, 11*. Available from: <https://commons.erau.edu/adfsl/2017/papers/11> [Accessed 30 April 2022].

Haraldseid, S., 2021. *Kan du stikke opp og gå gjennom databeslaget? Framgangsmåter for innholdsanalyse av databeslag og behovet for metodisk støtte*. Thesis (MSc). The Norwegian Police University College, Norway. Available from: <https://hdl.handle.net/11250/2760497> [Accessed 30 April 2022].

Haraway, D., 1991. *Simions, cyborgs and women: The reinvention of nature*. New York: Routledge.

Hewling, M.O., 2013. *Digital forensics: An integrated approach for the investigation of cyber/computer related crime*. Thesis (PhD). University of Bedfordshire, UK.

Horsman, G. and Sunde, N., 2020. Part 1: The need for peer review in digital forensics. *Forensic science international: digital investigation*, 35, 301062. DOI: 10.1016/j.fsidi.2020.301062.

Jahren, J.H., 2020. *Is the quality assurance in digital forensic work in the Norwegian police adequate?* Thesis (MSc). Norwegian University of Science and Technology, Norway. Available from: <https://hdl.handle.net/11250/2781174>.

James, J.I. and Gladyshev, P., 2013. A survey of digital forensic investigator decision processes and measurement of decisions based on enhanced preview. *Digital investigation*, 10(2), 148–157. DOI: 10.1016/j.diin.2013.04.005.

Kaufmann, M., 2017. The co-construction of crime predictions: Dynamics between digital data, software and human beings. In: N.R. Fyfe, H.O.I. Gundhus, and K.V. Rønn, eds. *Moral issues in intelligence-led policing*. London: Routledge, 143–160.

Knorr-Cetina, K.D., 1982. *The manufacture of scientific knowledge: An essay on the constructivist and contextual nature of science*. Oxford: Pergamon Press.

Knorr-Cetina, K.D., 1999. *Epistemic cultures*. Cambridge: Harvard University Press.

Kruse, C., 2016. *The social life of forensic evidence*. Oakland: University of California Press.

Kruse, C., 2021. Attaining the stable movement of knowledge objects through the Swedish criminal justice system: Thinking with infrastructure. *Science & technology studies*, 34(1), 2–18. DOI: 10.23987/sts.80295.

Lam, A., 2016. Making crime messy. In: D.R. Robert and M. Dufresne, eds. *Actor-Network Theory and crime studies*. London: Routledge, 51–66.

Latour, B., 2005. *Reassembling the social. An introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Latour, B. and Woolgar, S., 1979. *Laboratory life: The construction of scientific facts*. Thousand Oaks: Sage.

Marx, G.T. and Guzik, K., 2017. The uncertainty principle: Qualification, contingency, and fluidity in technology and social control. *In: M.R. McGuire and T.J. Holt, eds. The Routledge handbook of technology, crime and justice*. London: Routledge, 481–502.

McKemmish, R., 2008. When is digital evidence forensically sound? *In: I. Ray and S. Sheinoi, eds. Advances in digital forensics IV. Digital forensics 2008*. IFIP — The International Federation for Information Processing (Vol. 285). Springer. DOI: 10.1007/978-0-387-84927-0_1, 3-15.

Mol, A., 2002. *The body multiple: Ontology in medical practice*. Durham: Duke University Press.

NPCC, 2020. *Digital forensics science strategy. National Police Chiefs' Council*. Retrieved from:
<https://www.npcc.police.uk/FreedomofInformation/Reportsreviewsandresponsestoconsultations.aspx> [Accessed 30 April 2022].

Offit, A., 2019. Storied justice: The narrative strategies of US federal prosecutors. *In: J. Fleetwood, L. Presser, S. Sandberg, and T. Ugelvik, eds. The Emerald handbook of narrative criminology*. Bingley: Emerald Publishing, 72–96.

Page, H., *et al.*, 2019. A review of quality procedures in the UK forensic sciences: What can the field of digital forensics learn? *Science & justice*, 59(1), 83–92. DOI: 10.1016/j.scijus.2018.09.006.

Pennington, N. and Hastie, R., 1993. *The story model for juror decision making*. Cambridge: Cambridge University Press.

Rappert, B., Wheat, H., and Wilson-Kovacs, D., 2021. Rationing bytes: Managing demand for digital forensic examinations, *Policing and society*, 31(1), 52–65. DOI: 10.1080/10439463.2020.1788026.

Reedy, P., 2020. *Strategic leadership in digital evidence: What executives need to know*. Amsterdam: Academic Press.

Santos, F., 2014. Making sense of the story—The dialogues between the police and forensic laboratories in the construction of DNA evidence. *New genetics and society*, 33(2), 181–203. DOI: 10.1080/14636778.2014.916186.

Smit, N.M., Morgan, R.M., and Lagnado, D.A., 2018. A systematic analysis of misleading evidence in unsafe rulings in England and Wales. *Science & justice*, 58(2), 128–137. DOI: 10.1016/j.scijus.2017.09.005.

Stoykova, R., *et al.*, 2022. Reliability assessment of digital forensic investigations in the Norwegian police. *Forensic science international: Digital investigation*, 40, 301351. DOI: 10.1016/j.fsidi.2022.301351.

Stubbins, I.D. and Stubbins, N., 2008. On the horns of a narrative: Judgment, heuristics, and biases in criminal investigation. *In: K. Rossmo, ed. Criminal investigative failures*. London: Routledge, 123–164.

Sunde, N., 2021. What does a digital forensics opinion look like? A comparative study of digital forensics and forensic science reporting practices. *Science & justice*, 61(5) 586-596. DOI: 10.1016/j.scijus.2021.06.010.

Sunde, N., 2022. Strategies for safeguarding examiner objectivity and evidence reliability during digital forensic investigations. *Forensic science international: digital investigation*, 40, 301317. DOI: 10.1016/j.fsidi.2021.301317.

Sunde, N. and Dror, I.E., 2019. Cognitive and human factors in digital forensics: Problems, challenges, and the way forward. *Digital investigation*, 29, 101-108. DOI: 10.1016/j.diin.2019.03.011.

Sunde, N. and Dror, I.E., 2021. A hierarchy of expert performance (HEP) applied to digital forensics: Reliability and biasability in digital forensics decision making. *Forensic science international: digital investigation*, 37, 301175. DOI: 10.1016/j.fsidi.2021.301175.

SWGDE, 2018. *SWGDE Establishing confidence in digital and multimedia evidence forensic results by error mitigation analysis. Version: 2.0*, Scientific Working Group on Digital Evidence (November 20, 2018). Available from: <https://www.swgde.org/documents/Current%20Documents/SWGDE%20Establishing%20Confidence%20in%20Digital%20Forensic%20Results%20by%20Error%20Mitigation%20> [Accessed 30 April 2022].

Søe, S.O., 2019. A unified account of information, misinformation, and disinformation. *Synthese*, 198(6), 5929–5949. DOI: 10.1007/s11229-019-02444-x.

Tecuci, G., et al., 2016. *Intelligence analysis as discovery of evidence, hypotheses, and arguments: Connecting the dots*. Cambridge: Cambridge University Press.

The Law Commission, 1997. *Evidence in criminal proceedings: Hearsay and related topics (CP No 138, 1997) para 13.13*. The Law Commission, United Kingdom.

Tully, G., et al., 2020. Quality standards for digital forensics: Learning from experience in England & Wales. *Forensic science international: digital investigation*, 32, 200905. DOI: 10.1016/j.fsidi.2020.200905.

Van Koppen, P.J. and Mackor, A.R., 2020. A scenario approach to the Simonshaven case. *Topics in cognitive science*, 12(4), 1132–1151. DOI: 10.1111/tops.12429.

Wilson-Kovacs, D., 2019. Effective resource management in digital forensics: An exploratory analysis of triage practices in four English constabularies. *Policing: An international journal*, 43(1), 77–90. DOI: 10.1108/PIJPSM-07-2019-0126.

Wilson-Kovacs, D., 2021. Digital media investigators: Challenges and opportunities in the use of digital forensics in police investigations in England and Wales. *Policing: An international journal*, 44(4), 669–682. DOI: 10.1108/PIJPSM-02-2021-0019.

Article 4: Sunde, N. (2021). Strategies for safeguarding examiner objectivity and evidence reliability during digital forensic work. *Forensic Science International: Digital Investigation*, 40, 301317. <https://doi.org/10.1016/j.fsidi.2021.301317>.



Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

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

Strategies for safeguarding examiner objectivity and evidence reliability during digital forensic investigations

Nina Sunde

University of Oslo, The Norwegian Police University College, Pb. 2109 Vika, 0125, Oslo, Norway



ARTICLE INFO

Article history:

Received 24 August 2021

Received in revised form

27 November 2021

Accepted 30 November 2021

Keywords:

Digital forensics

Examiner objectivity

Evidence reliability

Error mitigation

Human factors

ABSTRACT

The paper investigates how digital forensic (DF) practitioners approach examiner objectivity and evidence reliability during DF investigations. Fifty-three DF practitioners responded to a questionnaire concerning these issues after first receiving a scenario, contextual information, and a task description and then examining an evidence file. The survey showed that hypotheses played an important role in how the DF practitioners handled contextual information before the analysis and safeguarding objectivity during the analysis. Dual tool verification was the most frequently referred measure for controlling evidence reliability. The analysis of responses revealed three significant causes of concern. First, 45% started the analysis without an innocence hypothesis in mind. Second, 34% applied no techniques to maintain their objectivity during the analysis. Third, 38% did not use any techniques to examine or control evidence reliability. The paper provides insights into how DF investigations are carried out in practice. These insights are essential for the DF community to guide the development of procedures that safeguard fair investigations and effective error mitigation strategies.

© 2021 The Author. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

A digital device may contain relevant information for shedding light on the six critical questions (who – what – where – when – why – how) of every investigation (Cook, 2016), and has great potential for providing investigative advice on source, activity, and offence level issues during the pre-trial stage (Sunde, 2021). According to the Digital Forensic Strategy issued by the UK National Police Chiefs Council, over 90% of all recorded crimes have a digital element (NPCC, 2020). However, digital evidence and its construction process are prone to technical and non-technical errors, which may occur randomly or systematically. Technical errors may happen randomly due to, for example, system or processing errors, but may also occur systematically because of, for example, programming flaws in software (SWGDE, 2018). In terms of non-technical errors, task-irrelevant contextual information has shown to be a significant source of biased observations and decisions in forensic science, including digital forensics (Cooper and Meterko, 2019; Sunde and Dror, 2021). Forensic science disciplines with a high degree of interpretation and subjectivity involved in the forensic process have a higher risk of such errors.

However, research has shown that even disciplines that are regarded as 'objective', such as DNA analysis and toxicology, are prone to bias (Dror, 2020).

Errors should be viewed as a natural and unavoidable part of any process involving human decision-making. In the context of a criminal investigation, error mitigation is necessary to, on the one hand, prevent errors from happening and, on the other, to detect the errors that may happen despite preventive efforts. It is, however, argued that rather than highlighting the problem of error and a 'bad apple' blame culture, the forensic community should instead discuss the risky nature of forensic science processes and enable dialogue around the management of risk (Earwaker et al., 2020). Implementing systems or measures for error mitigation is vital for preventing miscarriages of justice and is also an opportunity to enhance quality through systematic improvement and organisational learning.

This paper seeks to advance the sparse knowledge about digital forensic (DF) investigative strategies and practices. It explores how the DF practitioners approach the task of analysing an evidence file, and more specifically, which techniques (if any) they use to maintain their objectivity and to examine or control evidence reliability during the analysis. Such knowledge is crucial for designing adequate error mitigation techniques and quality management procedures for the DF domain. It examines the following research

E-mail address: nina.sunde@phs.no.

question: *How do DF practitioners handle contextual information, approach examiner objectivity and evidence reliability during the analysis of an evidence file?*

The paper is structured as follows: First, an outline of relevant research on DF practice is presented. The terms examiner objectivity and evidence reliability are then described and discussed. Further, the methods used to analyse the DF practitioners responses are described, followed by a presentation and discussion of the results. Eventually, a conclusion is offered.

1.1. Research on DF practice

While a large proportion of the research within the DF domain is concerned with technology, tools and techniques, a few empirical studies have examined DF practice. These studies centre on procedures and processes such as acquisition tasks (Carlton, 2007; Hewling, 2013), preview and triage (Wilson-Kovacs 2019; Rappert et al., 2021). A few studies have focused on implementation and compliance with quality procedures (Jahren, 2020; Tully et al., 2020). Others have explored the collaboration between investigating officers and DF practitioners (Hansen et al., 2017; Sunde, 2017) or tensions and professional dynamics between Digital Media Investigators, DF practitioners and investigating officers (Wilson-Kovacs, 2021). Brookman and Jones, (2021) studied practices concerning the construction of CCTV evidence in homicide investigations.

A few studies have focused on decision-making in DF investigations. James and Gladyshev (2013) conducted a small experimental study involving five DF practitioners where they examined decisions about exhibit exclusion during the enhanced preview of evidence files. The results showed that the participants were inconsistent in their decisions on which exhibits to include or exclude. Haraldseid (2021) interviewed six DF practitioners about strategies for the content analysis of evidence files and found that the applied approaches seemed to lack structure and that the capabilities of the tools influenced their decisions. For example, some chose not to perform keyword searches due to inefficiencies with the analysis software. Both studies suggested variability in decision-making during DF investigations. However, the generalisability from these studies is limited due to the small sample sizes.

Sunde and Dror (2021) conducted an experiment (further referred to as 'the DF experiment') to explore DF practice and decision-making. The current paper's sample and material originate from the DF experiment, and the context is therefore explained in detail below. The DF experiment primarily aimed to explore DF practitioners' biasability and reliability (consistency) according to the Hierarchy of Expert Performance (Sunde and Dror, 2021). Fifty-three DF practitioners from eight countries analysed the same evidence file obtained from the project Digital Corpora (Garfinkel et al., 2009). They read a short scenario description and were given the same assignment: 'What has happened, and what is Jean's involvement in the reported incident?'. The participants solved the task individually from their regular workstations and used the hardware and software of their choice. They spent four to five hours on the task, which included analysis and report writing.

To explore whether the participants were biased by contextual information, they were divided into four groups. The control group received the basic scenario, while the other three groups received variations of additional contextual information indicating guilt: a) Jean had confessed ('strong guilt' context); b), Jean could have participated by helping an insider to leak the information due to an ongoing wage conflict ('weak guilt' context); or innocence: c) Jean was framed in a phishing attack ('innocence' context) (see the full description in Appendix 1 in Sunde and Dror, 2021). The context significantly affected the observations of traces, but no biasing

effect was found on the interpretations or conclusions (Sunde and Dror, 2021). The reliability (consistency) between DF practitioners who received the same context was explored and revealed low reliability within all groups and at all examined levels (Sunde and Dror, 2021).

The reports from the DF experiment that included a conclusion ($n = 40$) were subject of a comparative combined qualitative and quantitative analysis aimed to examine DF reporting practices. The study explored the use of conclusion types, addressed issues (source, activity, offence), content relevant to the quality of the procedures/the result, and the use of certainty expressions (Sunde, 2021). The study found that DF practitioners tend to use Categorical conclusions and Strength of support (SoS) conclusion types and that the conclusions addressed source, activity, and offence levels. None of the reported conclusions referred to more than one hypothesis or scenario. They used a high variation of certainty expressions but did not explain their meaning or refer to a standardised framework. The study uncovered deficiencies in DF reporting practices, such as missing descriptions of the practitioners' competence or qualification, missing explication of complex terminology, vague or missing descriptions of the applied methods and tools, and the validity and reliability of these. The results were compared with a study of eight forensic science disciplines conducted by Bali et al. (2019, 2021), which showed that several of the identified deficiencies in DF reporting practices are shared with other forensic science disciplines.

1.2. Examiner objectivity and evidence reliability

The DF practitioners and their organisations should implement effective investigative strategies to manage contextual information, maintain examiner objectivity, and control evidence credibility. However, there is little research on DF practices concerning these issues, which is crucial for designing effective quality measures, and this paper seeks to expand the knowledge about these practices.

Objectivity is a cornerstone of forensic analysis aimed to provide legal decision-makers with unbiased and accurate results (Casey, 2011). The forensic analyst should search for incriminating and exculpating information and report their findings as thoroughly and impartially as possible to safeguard the rule of law. Several standards and guidelines for the DF discipline highlights the objectivity requirement (e.g., ACPO, 2012; ENFSI, 2015b; ISO, 2015; Jones et al., 2017; Interpol, 2019). Fulfilling the normative objectivity requirement may, however, be challenging for several reasons. DF work is sometimes carried out in close collaboration with the investigation team (Hansen et al., 2017; Sunde, 2017; Jones et al., 2020; Haraldseid, 2021). Such collaboration has advantages in obtaining timely results, for example by conducting a preliminary analysis of digital evidence before a suspect interview. It may also increase the DF practitioner's motivation, due to the feeling of meaningfulness and being part of something important (Sunde, 2017). A disadvantage of close cooperation is the risk of 'forensic confirmation bias' (Kassin et al., 2013) due to the exposure of task-irrelevant contextual information (Sunde and Dror, 2019, 2021). National bodies in the UK and USA have highlighted the risk of bias in forensic science and provided guidance stating that forensic analyses should be based upon task-relevant information (National Commission on Forensic Science, 2015; The Forensic Science Regulator, 2015).

Although it seems generally accepted that task-irrelevant information should be kept away from the forensic analyst, the principle is not straightforward to operationalise in practice, at least for some forensic disciplines. The first problem is concerned with deciding what is task-relevant and not. Due to the diverse nature or scope of a DF analysis assignment, what is clearly task-

irrelevant contextual information in one case may be crucial information in another. For example, knowing that the suspect confessed to the crime will most often be considered task-irrelevant information. However, if the DF practitioner is asked to verify information from the suspect's account to eliminate a false confession, the disclosure of details from the suspect's account without simultaneously revealing that the suspect confessed may sometimes be impossible. What is task-irrelevant context should, therefore, be critically evaluated and decided on a case-by-case basis.

The second problem is the procedure of excluding task-irrelevant information while ensuring access to the task-relevant information. This procedure may be straightforward for forensic disciplines primarily tasked with source evaluations (e.g., fingerprints, toxicology), while more complicated with those who work with activity (what activity/event happened) and even offence level issues (what crime – what level of guilt). Within forensic science, this challenge has been met with procedures for timing the introduction of contextual information during the evaluation process (Dror et al., 2015; Dror and Kukucka, 2021). Due to the controlled linear procedure, the initial examination of the evidence is unbiased by the contextual information, and the initial conclusion may later be adjusted to a certain extent after the introduction of the contextual information. To the author's knowledge, such procedures are not yet implemented in DF casework. A survey of how DF tasks were commissioned revealed a substantial risk that task-irrelevant information unintentionally could be forwarded to the DF practitioner because of how such procedures are designed and practiced (Sunde and Dror, 2021). The survey found that many submission forms encouraged the client to include case information but had no guidance on what to leave out. Several DF units used the submission forms as formal requests, but defined the scope of the task through dialogue with the client. The dialogue approach may be perceived as efficient due to the opportunity to explain and resolve ambiguities about the assignment. At the same time, dialogue poses a risk of providing task-irrelevant information due to the informal nature of such conversations.

The DF experiment (Sunde and Dror, 2021) described in Section 1.1 showed that DF practitioners who were given contextual information not directly relevant to the case under investigation were biased regarding the amount of traces they observed and deemed relevant to report (Sunde and Dror, 2021). However, contextual information is only one of several identified sources of bias in forensic work (Dror, 2020). The data set that is analysed may also challenge the examiner's objectivity (Dror, 2020). An evidence file from a smartphone will often expose the DF practitioner to a vast amount of task-irrelevant information, such as images, communication, and web search history, which may bias the examiner's observations and decisions. Regardless of how well the task-irrelevant contextual information is handled *before* the analysis, the practitioner must apply adequate techniques to maintain their objectivity *during* the analysis to minimise bias caused by the data itself. It is suggested that a hypothesis or proposition-based approach to the DF investigation may minimise bias (ENFSI, 2015a, 2015b; Sunde and Dror, 2019), but only a few small-scaled studies have explored how this is carried out in practice (Sunde, 2017; Haraldseid, 2021).

During DF casework in a criminal investigation, the practitioner should aim to produce relevant and credible results. The quality of the results depends on both technical and human factors. Evidence credibility includes the three components *authenticity*, *accuracy*, and *reliability*, which have implications for the evidential value (Anderson et al., 2005). Reliability is defined differently depending on context. For example, in a scientific context, Hand (2010) describes reliability as how different observers measure or score the

same phenomenon differently. ISO 27037:2012 describes reliability as 'to ensure digital evidence is what it purports to be' (ISO, 2012 p. 6). At the same time, Anderson et al. (2005) relate reliability to the devices used to produce the evidence and whether the applied process is repeatable, dependable, and consistent. It is generally accepted that DF tools may produce flawed results, and that error mitigation is necessary (SWGDE, 2018). Techniques such as 'dual tool' verification for controlling tool interpretation or hash sum calculation to ensure evidence integrity are often highlighted within DF authoritative literature (e.g., Casey, 2011; Flaglien, 2018), but there is little knowledge about which techniques are used in practice, how they are applied and documented.

2. Method

The empirical survey data for this paper were collected during the DF experiment (Sunde and Dror, 2021), which was outlined in Section 1.1. After completing the experiment and submitting their results, the DF practitioners completed a questionnaire ('Part 2 survey') with nine questions concerning how they approached the analysis and how they evaluated their findings. This paper is based on their responses to three open-ended questions from the questionnaire. In the first question, they were asked what they thought had happened after being informed about the scenario and before they started exploring the evidence file. In the second question, they were asked if they used any techniques to maintain objectivity during the analysis. In the third question, they were asked if they used any techniques to examine or control the evidence reliability during the analysis. Their responses were analysed in a combined qualitative-quantitative approach.

When exploring the responses in the first question (how they handled the context), it became apparent that most of the participants mentioned scenarios or hypotheses. A quantitative approach (performed in MS Excel Office Professional Plus 2016) was applied to examine the proportion that used hypotheses, the number of hypotheses they had developed, and how many that included an innocence hypothesis.

When analysing the responses concerning techniques for examiner objectivity and evidence reliability, they were first reviewed for whether they mentioned a technique or not. The responses that referred to techniques were coded for which techniques they mentioned, and similar or related techniques were assembled in the same code group. The analysis resulted in eight code groups for examiner objectivity techniques (none, hypotheses, focus on facts, avoidance, information requirement, forensic procedures, neutral presentation, other), and seven code groups for evidence reliability techniques (none, dual tool verification, metadata examination, cross-check of findings, hash calculation, manual verification of output, timeline analysis). When appropriate, quotes from the participants are included and annotated with the participant number, for example (P45). Quotes in the Norwegian language were translated to English by the author of the paper. The demographics of the participants are presented in Appendix 1.

3. Results and discussion

The results of the analysis are presented and discussed below. First, the results concerning the initial approach and belief after receiving the scenario/context are presented, followed by the applied techniques to maintain objectivity during the examination and how evidence reliability was examined and controlled. The results are discussed in the context of the findings from the former papers concerning the DF experiment (Sunde and Dror, 2021; Sunde, 2021) and other relevant research papers.

3.1. Initial approach – believe nothing or believe everything

To obtain knowledge about their initial beliefs before analysing the evidence file, the DF practitioners were asked the following question in the survey: ‘After reading the introduction about the case, and prior to the analysis – what did you think had happened in relation to the reported incident?’. The qualitative analysis showed that hypotheses, scenarios, or explanations were recurring themes in the responses. This finding is not very surprising since they (except the control group) had received contextual information indicating guilt or innocence for the suspect in addition to the basic background scenario description. However, more interesting was how many hypotheses the participants said they had in mind before they started the analysis.

The analysis showed that 10 (19%) participants replied that they did not have any opinion before the analysis. 19 (36%) participants stated that they believed a particular scenario had occurred, and 24 (45%) responded that they had developed two or more hypotheses, see Fig. 1. Among the latter, the number of hypotheses varied from two to five.

Among those who did not have an opinion, the analysis found two distinct approaches, which may be conceptualised as *active* and *passive mind-sets*. The passive approach was characterised by statements that they did not have an opinion. For example, ‘I did not have any particular opinion about what had happened, since I only had information from Alison’s incident report’ (P57). The active approach was characterised by actively and deliberately trying to suppress any hypotheses, for example: ‘I did not have an opinion, as a forensic research should’ (P58), or ‘I try to enter each examination with an open mind, scenario or not. I had no strong feelings about what had taken place and instead set my mind on the best approach to net the best result’ (P61).

The active avoidance of hypotheses may originate from the fallacy of believing that one is not biased, the so-called ‘bias blind spot’ (Dror, 2020; Pronin et al., 2002), or the ‘illusion of control’, which is a belief that one can overcome a bias with mere willpower (Dror, 2020).

Fig. 1 illustrates an important aspect of the results. The DF experiment (Sunde and Dror, 2021) showed that contextual information indicating guilt or innocence for the CFO Jean Jones tended to bias the DF practitioners’ observations. Fig. 1 shows that of those who replied that they had a single hypothesis in mind before the analysis, the majority had received either ‘strong guilt’ context

(suspect had confessed to the crime) or ‘innocence’ context (the suspect was framed in a phishing attack). Of those who had two or more hypotheses, the majority received either no context (the control group) or the ambiguous ‘weak guilt’ context (wage conflict in the firm). This finding illustrates *how* the contextual information influences the practitioners’ hypothesis development before the analysis, which in the DF experiment lead to biased observations. The ‘innocence’ and ‘strong guilt’ contexts lead to a one-sided starting point of the analysis for many of the participants. In contrast, those who received no context or the ambiguous ‘weak guilt’ context tended to start the analysis with more scenarios or hypotheses in mind.

The high proportion (36%) of participants with only one scenario in mind when starting the examination is of concern due to the risk of confirmation bias (Nickerson, 1998). Confirmation bias is characterised by a tendency to look for information that confirms the favoured hypothesis and overlook or explain away information that contradicts it. Due to satisficing strategies (Kahneman, 2011), the consequence may be that the examination is concluded too early and that relevant information is not uncovered. The DF experiment showed that the observations were biased and, for example, that those who had received context indicating that the suspect was innocent found less evidence than the other groups (Sunde and Dror, 2021).

The fact that almost half (45%) of the participants considered multiple hypotheses indicate that many were aware of the risk of biased and one-sided examinations. However, the *number* of hypotheses does not indicate whether they facilitate a fair investigation safeguarding the presumption of innocence (European Convention of Human Rights, article 6(2)). As stated in ENFSI (2015a), if ‘the alternative proposition’ is absent, the forensic practitioner may define a proposition that most likely reflects the party’s position. To operationalise the presumption of innocence, the proposition should reflect the party’s innocence – regardless of which level of issues the forensic questions address. For example, at the source level, the propositions could be: *Jean was the only one using the email address jean@m57.com vs the email address jean@m57.com was used by someone other than Jean*, and at activity level: *Jean sent the information outside M57.biz vs Jean did not send the information outside M57.biz*. Due to its relevance for maintaining examiner objectivity and safeguarding the presumption of innocence, the proportion that included an innocence hypothesis was explored.

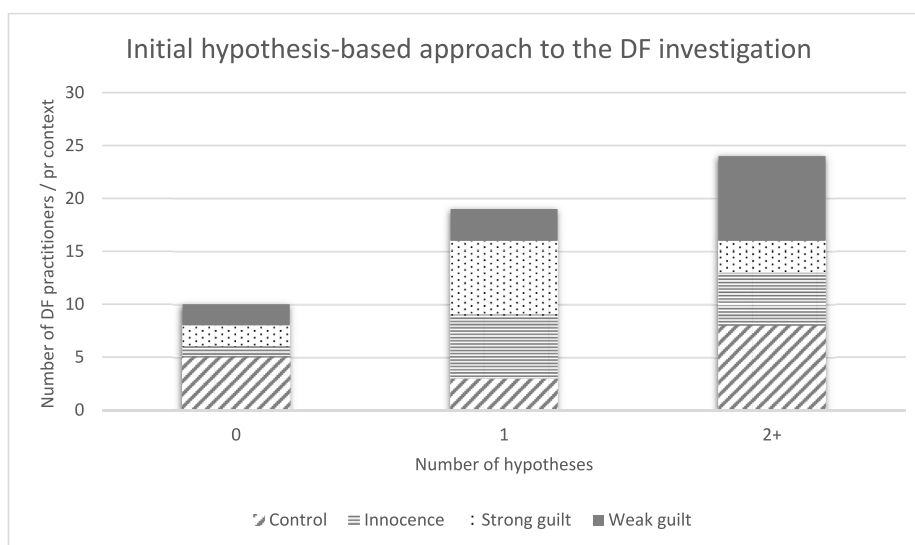


Fig. 1. Number of hypotheses prior to starting the analysis according to the participant survey.

Among those who referred to hypotheses, 29 (55%) had included an innocence hypothesis. More specifically, 7 of the 19 participants with a single hypothesis in mind before starting the analysis had framed it as an innocence hypothesis (Control: 0, Strong guilt: 1, Innocence: 5, Weak guilt: 1). Among the 24 participants who had developed two or more hypotheses, 22 had included an innocence hypothesis (Control: 7, Strong guilt: 3, Innocence: 5, Weak guilt: 7). This finding shows that those who had multiple hypotheses in mind before starting the analysis were more likely to include an innocence hypothesis than those who had none or a single hypothesis.

3.2. Maintaining examiner objectivity

To examine how they handled their objectivity during the analysis of the evidence file, the DF practitioners were asked the following question in the survey: 'Did you use any techniques to safeguard your objectivity during the analysis? If yes, please specify.'

Of the 53 participants, 18 (34%) replied that they did not use any techniques. Fig. 2 presents the identified approaches among the 35 participants who stated that they used a technique. The total number of techniques is higher than the number of respondents since some mentioned more than one technique.

17 participants mentioned approaches involving hypotheses or alternative explanations. 16 of these expressed that they applied the technique as a cognitive approach - to think about hypotheses - whilst only one participant stated to have written the hypotheses down. Several described that they focused on the facts/data. However, only one specified what to keep out of focus: 'Collate the necessary artefacts and review them in isolation from the case background' (P62). Some highlighted what they avoided during the examination, such as avoiding concluding too early, avoiding looking for guilt, or avoiding making invalid source-level conclusions. A few defined information requirements and only one explicitly stated that they were written down. Represented by the 'other' category, two commented that they used control procedures and checked the findings. One stated that he/she focused on the objective of the investigation, and one described using a mind map.

The large proportion of participants stating to have used hypotheses may signify the operationalisation of the the presumption

of innocence principle, where the DF practitioner is obliged to look for both incriminating and exculpating evidence. The use of hypotheses may also indicate that DF practitioners apply the scientific method to their analysis, where hypotheses play a central part (see, e.g. Platt, 1964). It may also point towards an awareness concerning the risk of cognitive bias. However, the low number of participants who stated that they would write the hypotheses (and their related information requirements) down indicates an overconfidence concerning their ability to systematically test the hypotheses as a mental activity. This finding corresponds with the results concerning biasability in the DF experiment, which showed that the participants were biased at the observation level, namely how many traces they found and deemed relevant to describe in their reports. Together, these findings indicate that merely thinking about hypotheses is an insufficient measure to avoid bias. Instead, writing the hypotheses down and systematically testing them would probably be a more effective bias minimisation strategy. This is supported by Rassin (2018)'s research, who found that a pen-and-paper tool reduced tunnel vision when evaluating whether the same evidence supported different hypotheses.

The 'avoidance' technique may result from the 'illusion of control' (Thornton, 2010), which is one of several fallacies about bias commonly held by experts (Dror, 2020). Instead of reducing bias, avoidance may have the opposite effect and increase the effect of the bias (Dror, 2020). Wallace (2015, p. 65) emphasises focusing on facts in the investigation to avoid one-sided 'case building' where a guilt hypothesis is driving the search for information. However, it may be argued that 'focus on facts' also shares characteristics with the 'avoidance' technique since it implicitly states that something should be kept out of the focus or avoided.

The large proportion that did not apply any technique may relate to the 'bias blind spot' (Pronin et al., 2002) or the fallacy of 'expert immunity' (Dror, 2020), where one believes that expertise is adequate protection against bias.

3.3. Safeguarding evidence reliability

To obtain knowledge about how they handled the evidence reliability, the DF practitioners were asked the following question in the survey: 'Did you use any techniques to examine or control the reliability of the evidence during the analysis? If yes, please specify.'

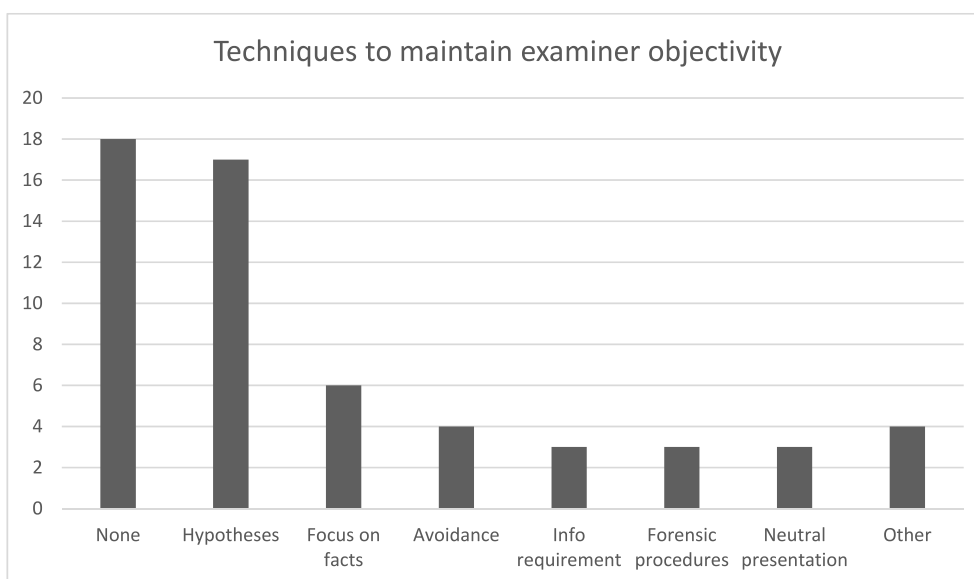


Fig. 2. Techniques used to maintain objectivity during the DF investigation.

As described, reliability may be associated with various meanings. The term was not defined or explicated in the survey. Therefore, all reported techniques were included regardless of whether they fit into the broad credibility definition or the narrower reliability definition outlined in Section 1.2.

Of the 53 participants, 20 (38%) replied that they did not use any techniques to examine or control the evidence reliability, and 4 of these justified it with time constraints. Among the 33 participants who responded to have used one or more techniques, six approaches were identified (see Fig. 3).

Some of the reported techniques were useful for checking the reliability (consistency) of the tool interpretation. An expert may have the knowledge and skills to conduct such a verification manually. An alternative approach is 'dual tool' verification (Turner, 2005), where two different tools are used to examine the data and check for variations in the tool's interpretations. In the current study 'dual tool' verification was mentioned by 13 participants, and was the most frequently referred approach. The technique is recommended in many guidelines and standards (e.g., Interpol, 2019; Jones et al., 2017; ACPO 2011; ACPO 2012) and can be an efficient way of determining whether there are reasons for checking the results more closely. However, the approach has been criticized for its numerous limitations (Friheim, 2016; Nordvik et al., 2021). The concept rests on the assumption that if two different tools interpret the data similarly, they are more likely to be valid. The validity of this assumption depends on whether the tools are built upon different libraries, engines, and methods (Friheim, 2016). Many of the DF investigation tools have secret source code, which complicates the testing of these assumptions. Using two tools that share libraries, engines or methods may, in the worst-case, result in a similar but flawed interpretation of the same data and create an illusion of valid results. These limitations imply that although using 'dual tool' verification may seem straightforward, knowing which tools to use and evaluating the strength of the result requires more advanced knowledge and skills. Accurate documentation of which tools were used to verify results is thus crucial for the transparency and evaluation of the result's credibility.

Several of the reported techniques, such as metadata examination and hash calculations, were aimed to control the authenticity or accuracy of the findings. The use of hash calculation is a tool-

based technique to check whether two files/datasets are identical, resulting in a categorical 'match' or 'no match' conclusion. The reported metadata examinations were either related to files ($n = 8$) or email headers ($n = 3$). The examination and interpretation of metadata in files or mail headers would require more expertise. Some used investigative techniques, such as cross-checking the findings or using a timeline.

All the applied techniques have in common that they aim to detect technical flaws. None stated to have used techniques for uncovering human errors, such as biased examinations or one-sided or overstated conclusions. Peer review or audit of the report and verification of findings by a second practitioner are examples of such measures (Horsman and Sunde, 2020). Only one participant mentioned that peer review would be carried out under other circumstances: 'Normally, [unit] procedures require an audit of a second researcher. Due to limited time, this was not carried out now' (P58).

The large number of participants that did not use any technique adds new and valuable information to the analysis of reports from the DF experiment, which showed that many of the reports lacked detailed descriptions methods and tools used for the analysis, as well as the reliability and validity of these (Sunde, 2021). Without this information, the different actors in the judicial system have minimal opportunities to make an informed assessment about the quality of the evidence and the process where it was produced and are sometimes left with the choice of whether one should trust the expert or not. However, the participants' responses to the questionnaire indicate that the problem is not just related to inadequate transparency and documentation practices, but also insufficient quality control and error mitigation practices.

3.4. Limitations and future directions

The sample of participants and material obtained through the DF experiment are considered adequate and representative of current DF practice. The participants were all DF practitioners, with a variation in gender, organisational level, educational background, competence, and experience with DF work (see Appendix 1). The DF experiment had high ecological validity due to an experimental design that mimicked regular DF casework: They received some

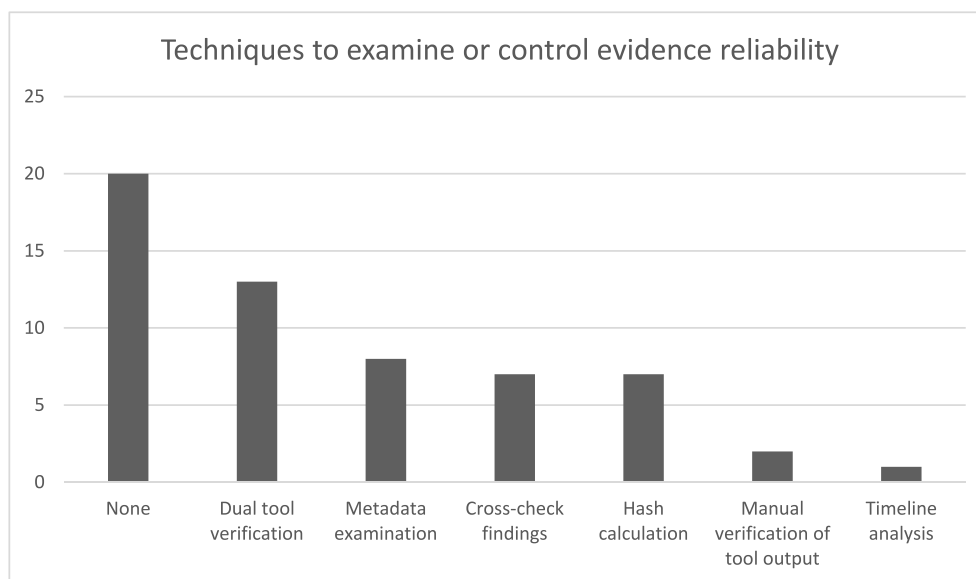


Fig. 3. Techniques used to examine or control evidence reliability during the DF investigation.

background information about the case, an evidence file, and a task description. They processed and analysed the evidence file with tools and methods they would typically use and were encouraged to use their standard structure or template for the report. The survey thus referred to a situation that was similar to regular DF investigations.

Some possible limitations should be discussed. First, 53 participants might be considered a limited sample for a quantitative study such as a questionnaire. However, the sample is quite large for the type of experiment that formed the basis for the participants' reflections when completing the survey. A systematic review of cognitive bias research in forensic science showed that experimental studies involving practitioners (as opposed to students) often involved less than 15 participants and typically varied from 6 to 70 participants, with a few exceptions (Cooper and Meterko, 2019). Due to limited knowledge about the DF investigator population, a sampling strategy involving voluntary participation, and a possible 'self-selection bias' (see, for example, Heckman, 2010), it is uncertain how representative the participant sample is for the total DF investigator population. The sample described in Appendix 1 should therefore be assessed when considering the transferability of the results to own organisation.

Second, the time limitation in the DF experiment (4–5 h) may have affected the number of techniques used for safeguarding examiner objectivity and evidence reliability, and consequently, the amount referred to in the survey. However, it should be noted that only four participants mentioned time constraints.

Third, the participants knew that they participated in the research. The research effect may thus have influenced their responses, for example, to describe techniques they ought to have used, rather than what they actually used during the analysis. On the other hand, some may have used techniques they did not associate with 'objectivity' or 'reliability' and therefore left them out in their responses.

Fourth, the participants mentioned various techniques for safeguarding examiner objectivity and evidence reliability, yet their responses do not expose whether they implemented the techniques correctly and effectively. For example, many stated that they had developed multiple hypotheses (including an innocence hypothesis). However, since they were not written down and systematically tested, they may have failed to minimise bias during the DF experiment. In addition, the thoroughness of metadata review and the level of expertise to perform a valid interpretation may vary between different practitioners. Therefore, future research should examine how these techniques are applied and performed in practice to gain more knowledge of whether they are effective for the intended purpose.

4. Conclusion

This paper contributes to the sparse but growing body of research on DF investigative practice. Specifically, it provides novel insights into how DF practitioners handle contextual information and safeguard examiner objectivity and evidence reliability during examinations of digital evidence, which are essential for understanding how human factors influence the quality of digital evidence. The findings are vital for countering erroneous beliefs about digital evidence per default being objective and reliable due to a construction process characterised by 'mechanical objectivity' (Daston, 1992; Daston and Galison, 2007). Insights into how DF investigations are carried out in practice are essential to the DF community to guide the development of analysis procedures safeguarding fair investigations and effective error mitigation strategies.

The results indicate a correspondence between the type of

context received and the number of hypotheses formed by the DF practitioners before the analysis. Most DF practitioners started the analysis with one or more hypotheses in mind about what had happened. A majority of those who had a single hypothesis in mind when starting the analysis had received the 'strong guilt' or 'innocence' context. In contrast, the majority of those who had developed two or more hypotheses had received no context (the control group) or the ambiguous 'weak guilt' context. DF practitioners who started the analysis with two or more hypotheses more frequently included an innocence hypothesis than those with only one hypothesis in mind.

Approximately a third of the DF practitioners replied that they did not use any technique for maintaining examiner objectivity during the analysis. Among those that stated to have used techniques, nine approaches were identified. The hypothesis-based approach was the most frequently referred technique. Previous research has suggested that a hypothesis/proposition based approach to DF investigations may minimise bias, but only a few qualitative studies have explored how this is carried out in practice (Sunde, 2017; Haraldseid, 2021). The results presented in this paper indicate that merely *thinking about* hypotheses is not an adequate bias minimising strategy and suggest that writing the hypotheses down and systematically testing them may be a better approach. To safeguard the presumption of innocence, at least one hypothesis should implicate innocence for the person suspected of a crime. However, more research is necessary for examining to what extent the suggested measures are effective for safeguarding high-quality investigations and a fair administration of justice.

As much as 38% replied that they did not use any technique to examine or control the results' reliability. Among those who stated to have used a technique, six approaches were identified where 'dual tool' verification was the most frequently referred technique. The technique is recommended in several best practice guidelines and standards, although there are several known limitations. Safely applying 'dual tool' verification requires expertise to select adequate tools and transparency concerning the applied procedure and results.

While much is written in and best practice guidelines, standards and academic literature about techniques to safeguard examiner objectivity and evidence reliability, this paper shows that there might be a gap between the recommendations and how this is handled in practice. Due to the risk of random and systematic errors of both technical and non-technical types, the large proportion that did not apply any techniques to safeguard examiner objectivity or evidence reliability is a cause for concern.

Declaration of competing interests

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.

Funding

The paper was funded by a grant from the Norwegian Police University College awarded to the author for pursuing a PhD.

Acknowledgements

I would like to thank Professor Helene O. I. Gundhus, Associate Professor Jon Strype, and the anonymous reviewers for insightful comments and constructive criticism of former versions of this paper.

Appendix 1 Demographics

Variable	Value	Control N = 16	Strong Guilt N = 12	Weak Guilt N = 13	Innocent N = 12
Country*	Norway	75	83	85	92
	Other	25	17	15	8
Gender*	Men N = 44	81	100	85	67
	Women N = 9	19	0	15	33
Educational background*	Civil	63	58	54	66
	Police/both	38	42	46	33
Educational level*	PhD	6	0	0	0
	MSc	25	33	23	33
	BSc	69	58	69	58
	Lower	0	8	8	8
Organisational level*	National	31	25	54	42
	Local	63	67	46	58
	Private industry	6	8	0	0
Mean years of experience within law enforcement	With crim. inv.	5	6	5	6
	With DF	5	4	3	4

*percentage distribution of background variables per group.

References

- ACPO, 2011. ACPO managers guide. Good practice and advice guide for managers of e-crime investigation, Version 0.1.4. Association of Chief Police Officers.
- ACPO, 2012. Good practice guide for digital evidence, Version 5. Association of Chief Police Officers.
- Anderson, T., Schum, D., Twining, W., 2005. Analysis of evidence. Second ed. Cambridge University Press, Cambridge.
- Bali, A.S., Edmond, G., Ballantyne, K.N., Kemp, R.I., Martire, K.A., 2019. Communicating forensic science opinion: An examination of expert reporting practices. *Sci. Justice* 60 (3), 216–224. <https://doi.org/10.1016/j.scijus.2019.12.005>.
- Bali, A.S., Edmond, G., Ballantyne, K.N., Kemp, R.I., Martire, K.A., 2021. Corrigendum to “Communicating forensic science opinion: An examination of expert reporting practices” [*Sci. Justice* 60 (3) (2020) 216–224]. *Sci. Justice* 61 (4), 449–450. <https://doi.org/10.1016/j.scijus.2021.04.001>.
- Brookman, F., Jones, H., 2021. Capturing killers: the construction of CCTV evidence during homicide investigations. *Polic. Soc.* 1–20. <https://doi.org/10.1080/10439463.2021.1879075>.
- Carlton, G.H., 2007. A grounded theory approach to identifying and measuring forensic data acquisition tasks. *J. Digit. Forensic Secur. Law* 2 (1), 35–55. <https://doi.org/10.15394/jdfsl.2007.1015>.
- Casey, E., 2011. Digital evidence and computer crime: Forensic science, computers, and the internet. Academic Press, Amsterdam.
- Cook, T., 2016. Blackstone's senior investigating officers' handbook, Fourth ed. Oxford University Press, Oxford.
- Cooper, G.S., Meterko, V., 2019. Cognitive bias research in forensic science: A systematic review. *Forensic Sci. Int.* 297, 35–46. <https://doi.org/10.1016/j.forsciint.2019.01.016>.
- Daston, L., 1992. Objectivity and the escape from perspective. *Soc. Stud. Sci.* 22 (4), 597–618. <https://doi.org/10.1177/030631292022004002>.
- Daston, L., Galison, P., 2007. Objectivity. Zone Books, New York.
- Dror, I.E., 2020. Cognitive and human factors in expert decision making: Six fallacies and the eight sources of bias. *Anal. Chem.* 92 (12), 7998–8004. <https://doi.org/10.1021/acs.analchem.0c00704>.
- Dror, I.E., Kukucka, J., 2021. Linear sequential unmasking—expanded (LSU-E): A general approach for improving decision making as well as minimizing noise and bias. *Forensic Science International: Synergy* 3, 100161. <https://doi.org/10.1016/j.fsisyn.2021.100161>.
- ENFSI, 2015a. ENFSI guideline for evaluative reporting in forensic science. Strengthening the evaluation of forensic results across Europe (STEOFRAE). Version 3.0. https://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf. (Accessed 15 December 2021).
- Dror, I.E., Thompson, W.C., Meissner, C.A., Kornfield, I., Krane, D., Saks, M., Risinger, M., 2015. Letter to the editor - context management toolbox: A linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *J. Forensic Sci.* 60 (4), 1111–1112. <https://doi.org/10.1111/1556-4029.12805>.
- Earwaker, H., Nakhaeizadeh, S., Smit, N.M., Morgan, R.M., 2020. A cultural change to enable improved decision-making in forensic science: A six phased approach. *Sci. Justice* 60 (1), 9–19. <https://doi.org/10.1016/j.scijus.2019.08.006>.
- ENFSI, 2015b. Best practice manual for the forensic examination of digital technology. ENFSI-BPM-FOT-01, Version 01, November 2015. https://enfsi.eu/wp-content/uploads/2016/09/1_forensic_examination_of_digital_technology_0.pdf. (Accessed 15 December 2021).
- Flaglien, A.O., 2018. The digital forensic process. In: Arnes, A. (Ed.), *Digital Forensics*. Wiley, Hoboken.
- Friheim, I., 2016. Practical use of dual tool verification in computer forensics. Master's Thesis. University College Dublin, Ireland.
- Garfinkel, S., Farrell, P., Roussev, V., Dinolt, D., 2009. Bringing science to digital forensics with standardized forensic corpora. DFRWS 2009, Montreal, Canada. <https://simson.net/clips/academic/2009.DFRWS.Corpora.pdf>. (Accessed 15 December 2021).
- Hand, D.J., 2010. Measurement theory and practice: the world through quantification. Wiley, Hoboken.
- Hansen, H.A., Andersen, S., Axelsson, S., Hopland, S., 2017. Case study: a new method for investigating crimes against children. Annual ADFSL Conference on digital forensics, Security and law, 11. Available at: <https://commons.erau.edu/adfsl/2017/papers/11>. (Accessed 15 December 2021).
- Haraldseid, S., 2021. Kan du stikke opp og gå gjennom databeslaget? Framgangsmåter for innholdsanalyse av databeslag og behovet for metodisk støtte. [Content analysis of seized data: 'Could you please go analyse all the digital evidence?'] Master's Thesis. The Norwegian Police University College, Norway.
- Heckman, J.J., 2010. Selection bias and self-selection. In: Durlauf, S.N., Blume, L.E. (Eds.), *Microeconometrics*. The New Palgrave Economics Collection. Palgrave Macmillan, London, pp. 242–266.
- Hewling, M.O., 2013. Digital Forensics: an Integrated approach for the investigation of cyber/computer related crimes. PhD Thesis. University of Bedfordshire, UK.
- Horsman, G., Sunde, N., 2020. Part 1: The need for peer review in digital forensics. *Forensic Sci. Int.: Digit. Invest.* 35, 301062. <https://doi.org/10.1016/j.fsi.2020.301062>.
- Interpol, 2019. Global guidelines for digital forensic laboratories. Available at: https://www.interpol.int/content/download/13501/file/INTERPOL_DFL_GlobalGuidelinesDigitalForensicsLaboratory.pdf. (Accessed 15 December 2021).
- ISO, 2015. Information technology - Security techniques - Guidelines for the analysis and interpretation of digital evidence. (ISO/IEC 27042:2015).
- ISO, 2012. Information technology - Security techniques - Guidelines for identification, collection, acquisition and preservation of digital evidence. (ISO/IEC 27037:2012).
- Jahren, J.H., 2020. Is the quality assurance in digital forensic work in the Norwegian police adequate? Master's Thesis. Norwegian University of Science and Technology, Norway.
- James, J.L., Gladyshev, P., 2013. A survey of digital forensic investigator decision processes and measurement of decisions based on enhanced preview. *Digit. Invest.* 10 (2), 148–157. <https://doi.org/10.1016/j.diin.2013.04.005>.
- Jones, H., Brookman, F., Williams, R., Fraser, J., 2020. We need to talk about dialogue: Accomplishing collaborative sensemaking in homicide investigations. *Police J.* 94 (4), 572–589. <https://doi.org/10.1177/0032258X20970999>.
- Jones, N., Voelzow, V., Bradley, A., Stamenkovic, B., 2017. Digital forensics. A basic guide for the management and procedures of a digital forensics laboratory. Council of Europe.. Version 1.1.
- Kahneman, D., 2011. *Thinking, fast and slow*. Macmillan, New York.
- Kassin, S., Dror, I.E., Kukucha, J., 2013. The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition* 2 (1), 42–52. <https://doi.org/10.1016/j.jarmac.2013.01.001>.
- National Commission on Forensic Science, 2015. Ensuring that forensic analysis is based upon task-relevant information. Human Factors Subcommittee. Retrieved from: <https://www.justice.gov/archives/ncfs/page/file/641676/download>. (Accessed 15 December 2021).
- Nickerson, R.S., 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* 2 (2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>.
- Nordvik, R., Stoykova, R., Franke, K., Axelsson, S., Toolan, F., 2021. Reliability validation for file system interpretation. *Forensic Sci. Int.: Digit. Invest.* 37, 301174.

- <https://doi.org/10.1016/j.fsidi.2021.301174>.
- NPCC, 2020. Digital forensics science strategy. National Police Chief Council. Retrieved from. <https://www.npcc.police.uk/FreedomofInformation/Reportsreviewsandresponsestoconsultations.aspx>. (Accessed 15 December 2021)
- Platt, J.R., 1964. Strong inference. *Science* 146 (3642), 347–353.
- Pronin, E., Lin, D.Y., Ross, L., 2002. The bias blind spot: Perceptions of bias in self versus others. *Pers. Soc. Psychol. Bull.* 28 (3), 369–381. <https://doi.org/10.1177/0146167202286008>.
- Rappert, B., Wheat, H., Wilson-Kovacs, D., 2021. Rationing bytes: managing demand for digital forensic examinations. *Polic. Soc.* 31 (1), 52–65. <https://doi.org/10.1080/10439463.2020.1788026>.
- Rassin, E., 2018. Reducing tunnel vision with a pen-and-paper tool for the weighting of criminal evidence. *J. Investigative Psychol. Offender Profiling* 15 (2), 227–233. <https://doi.org/10.1002/jip.1504>.
- Sunde, N., 2021. What does a digital forensic opinion look like? A comparative study of digital forensics and forensic science reporting practices. *Sci. Justice* 61 (5), 586–596. <https://doi.org/10.1016/j.scijus.2021.06.010>.
- Sunde, N., 2017. Non-technical sources of errors when handling digital evidence within a criminal investigation. Master's Thesis. Norwegian University of Science and Technology, Norway.
- Sunde, N., Dror, I.E., 2021. A Hierarchy of expert performance (HEP) applied to digital forensics: Reliability and biasability in digital forensics decision making. *Forensic Sci. Int.: Digit. Invest.* 37, 301175. <https://doi.org/10.1016/j.fsidi.2021.301175>.
- Sunde, N., Dror, I.E., 2019. Cognitive and human factors in digital forensics: Problems, challenges, and the way forward. *Digit. Invest.* 29, 101–108. <https://doi.org/10.1016/j.diin.2019.03.011>.
- SWGDE, 2018. SWGDE Establishing confidence in digital and multimedia forensic results by error mitigation analysis. Version: 2.0 (20 November 2018). Scientific Working Group on Digital Evidence.
- The Forensic Science Regulator, 2015. Cognitive bias effects relevant to forensic science examinations. Retrieved from. <https://www.gov.uk/government/publications/cognitive-bias-effects-relevant-toforensic-science-examinations>. (Accessed 15 December 2021).
- Thornton, J.L., 2010. A rejection of working blind as a cure for contextual bias. *J. Forensic Sci.* 55 (6), 1663. <https://doi.org/10.1111/j.1556-4029.2010.01497.x>.
- Tully, G., Cohen, N., Compton, D., Davies, G., Isbell, R., Watson, T., 2020. Quality standards for digital forensics: Learning from experience in England & Wales. *Forensic Sci. Int.: Digit. Invest.* 32, 200905. <https://doi.org/10.1016/j.fsidi.2020.200905>.
- Turner, P., 2005. Digital provenance – interpretation, verification and corroboration. *Digit. Invest.* 2 (1), 45–49. <https://doi.org/10.1016/j.diin.2005.01.002>.
- Wilson-Kovacs, D., 2021. Digital media investigators: challenges and opportunities in the use of digital forensics in police investigations in England and Wales. *Policing: Int. J.* 44 (4). <https://doi.org/10.1108/PJPSM-02-2021-0019>.
- Wallace, W.A., 2015. The effect of confirmation bias in criminal investigative decision making. PhD Thesis. Walden University, USA.
- Wilson-Kovacs, D., 2019. Effective resource management in digital forensics: An exploratory analysis of triage practices in four English constabularies. *Policing: Int. J.* 43 (1), 77–90. <https://doi.org/10.1108/PJPSM-07-2019-0126>.

Article 5: Sunde, N. (2021). What does a digital forensics opinion look like? A comparative study of digital forensics and forensic science reporting practices. *Science & Justice*, 61(5) 586-596. <https://doi.org/10.1016/j.scijus.2021.06.010>.



Original Research Article

What does a digital forensics opinion look like? A comparative study of digital forensics and forensic science reporting practices

Nina Sunde

Norwegian Police University College, Postboks 2109 Vika, 0125 Oslo, Norway

ARTICLE INFO

Keywords:

Digital Forensics
Forensic Science
Conclusion types
Opinion evidence
Certainty expressions
Reporting practices

ABSTRACT

This study explores digital forensics (DF) reporting practices and compares the results with other forensic science disciplines. Forty reports were obtained from a quasi-experiment involving DF examiners, and a quantitative content analysis was performed to determine which conclusion types they applied and which content they included with relevance to the credibility of the reported results. A qualitative analysis was performed to examine the certainty expressions used in the conclusions. The results were compared to a study of eight forensic science disciplines performed by Bali et al. [24,26]. The results show that the DF examiners tend to present their conclusions either as Categorical conclusion or Strength of support (SoS) conclusion types and that they address source, activity, and offence level issues in their conclusions. The content analysis indicates deficiencies in DF reporting practices, and several of the challenges seem to be shared with other FS disciplines. The analysis of certainty expressions showed that a plethora of expressions was used, and that they lacked reference to an established framework. The results indicate that more research on DF evaluation and reporting practices is necessary and justifies a need for enhanced focus on quality control and peer review within the DF discipline.

1. Introduction

Identifying, collecting, analysing, and presenting digital evidence to legal decision-makers often requires expert knowledge, and sometimes also highly specialised skills within the digital forensics (DF) discipline on particular technologies, techniques, or methods. The results of DF investigations are documented in reports and disseminated to legal decision-makers which often have little to no knowledge of or insight into the processes that have led to the digital evidence to which they are presented. The legal decision-makers rarely possess sufficient technical competence to scrutinize the quality of the digital evidence itself or the process in which it was produced. This is not unique to the DF discipline. The justice system relies on a wide range of forensic expert disciplines to analyse and present evidence to legal decision-makers, such as DNA evidence, fingerprint analysis, toxicology analysis, or forensic pathology.

The trust in experts and expert systems is of vital importance for the complex and fragmented modern society to be well functioning [1]. A challenge to the trust is the use of technology with secret source code and functions, often referred to as “black box” technology, which has been the subject of debate within the DF community [e.g., 2–6].

Although parts of the DF process may involve technology,

automation, and computation, the human instrument – the DF examiner – plays an active role in the process of producing digital evidence [7,8]. The digital evidence would probably be useless to the legal decision-makers if it was presented in its original form (zeroes and ones). To make sense as a piece of evidence, it passes through several instances of human and technology-based interpretation. The analysis report is the primary medium for presenting the evidence, and by using their expertise, the DF examiners construct the digital evidence by explaining its relevance, meaning, and credibility to the legal decision-makers. How the legal decision-makers understand these aspects depends on *what* information the DF examiners include or exclude, and *how* the information is articulated. At this point where different epistemic cultures meet, there is no guarantee that what the DF examiners intended to convey is what the legal decision-makers perceive when they read the reports or attend presentations of the result.

An important aspect of evaluating and presenting digital evidence is the distinction between facts and opinions. What is considered a fact depends on the epistemological perspective, which is different in law vs science. While law treats a fact as “an unproblematic, uncontested inferential conclusion” [9, p. 30], here facts should be understood as statements that are generally made and accepted as a matter of routine practice and convention and involve little to no active interpretation

E-mail address: Nina.Sunde@phs.no.

<https://doi.org/10.1016/j.scijus.2021.06.010>

Received 13 January 2021; Received in revised form 15 June 2021; Accepted 28 June 2021

Available online 1 July 2021

1355-0306/© 2021 The Author(s). Published by Elsevier B.V. on behalf of The Chartered Society of Forensic Sciences. This is an open access article under the CC

BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

from the expert [9]. An opinion would, in this context, be statements about evidence that involves interpretation or evaluation from the expert. Suppose DF examiners ignore or are unaware of the limitations and uncertainties of the digital evidence they have produced, and present it as facts. In that case, there is a risk of misleading the legal decision-makers, and in the worst-case lead to erroneous verdicts such as conviction of innocent people and acquittal of guilty culprits.

After evaluating 235 cases of rulings argued unsafe from the Court of Appeal of England and Wales, Smit et al. [10] found that flawed or misleading expert evidence had caused or contributed to errors of justice. They found that forensic evidence (including digital evidence) was the second most frequent evidence type (32%) in these cases and highlighted interpretation at the activity level as a significant problem [10]. Sunde and Dror [8] described the cognitive and human factors that could cause errors during the DF process, and a quasi-experiment with DF examiners showed that their observations of traces were prone to bias and that they were inconsistent in their observations of traces, interpretations, and conclusions [11].

It seems to be generally acknowledged that there is a need to develop a robust quality regime for the DF discipline [e.g., 8,12–16]. One of the fundamental issues in this respect is how DF opinion evidence should be documented and reported. There will always be a degree of uncertainty associated with an opinion, and to facilitate a fair assessment of evidence, the uncertainty should be disseminated clearly and understandably. As demonstrated by Georgiou et al. [17], a plethora of qualitative and quantitative approaches may be applied to evaluate and present uncertainty within forensic science.

An essential step of the DF maturation process was to bridge the DF discipline with the forensic science (FS) domain and harmonise practices by defining and describing core forensic processes [18]. One of the identified common processes was *evaluation*, which means to “Produce a value that can be fed into a decision process” [18, p. 8]. Pollitt et al. [18] relate evaluation to the Bayesian model of probability and inference. According to this model, trace evaluation should be performed relative to at least two competing claims as a “strength of evidence” approach as opposed to a “strength of hypothesis” approach. The evaluation results may sometimes be summarised as a likelihood ratio (LR) [18]. The Bayesian approach has been demonstrated with, for example, smartphone tampering, file recovery, deletion of evidence, backdating [19], location-related mobile device evidence [20], and cell site analysis [21].

The Bayesian model of probability and inference and the use of LR have gained ground within forensic science during the last decades and are regarded by many forensic scientists and mathematicians as the only rational means of evaluating and presenting scientific evidence [22,23]. The use of LR in forensic science is, however, still debated. Due to the complexity and diversity of digital traces, some argue that issues may not always be suitable for a formalised evaluation of LR; “the complexity and multi-faceted nature of uncertainty cannot always be reduced into a numerical formula and is not always easily quantifiable” [17, p. 314]. Very few FS disciplines (e.g., DNA analysis) have adequate databases and statistical modelling techniques in place to calculate the LR [24]. Due to the rapid and significant changes in information technology, developing and maintaining databases for DF traces would be a challenging, if not an impossible, task. Without such a foundation, the LR is produced based on the expert’s knowledge, experience, and experimental testing. It is also debated whether the legal decision could be misled to assign too much evidential weight to the evidence when presented to a mathematically precise LR quantification which in reality may have no rigorous empirical foundation and may conceal the sources of subjectivity rather than exposing them [24,25].

The ongoing normative discourse seems to lack a vital knowledge component: The practice of DF examiners. The current study aims to provide more knowledge about DF examiners’ practice when they interpret, evaluate, and present digital evidence in their reports. Due to the efforts aimed at harmonising DF and FS practices, the results are compared with a study of conclusions from several FS disciplines

conducted by Bali et al. [24,26].

The study explored the following research questions: First, what characterises the opinions used in DF reports in terms of opinion type, uncertainty expressions, addressed issues, and included content relevant to the credibility assessment of the reported results. Second, to what degree does the DF reporting practice concerning applied opinion types and included content deviate from other FS disciplines.

The article is structured as follows: First, the methods used in this study are presented, which include a description of the sample and material, followed by an elaboration of the quantitative analysis of conclusions and report content and the qualitative analysis of certainty expressions. The results are then presented, analysed, and discussed, followed by a summary discussion where the study’s limitations are outlined and debated. Eventually, a conclusion is presented.

2. Method

2.1. Sample and material

The current study included a sample of reports collected in a DF practice and decision-making study conducted from September 2019 to January 2020 [11]. DF examiners from eight different countries participated in a quasi-experiment (further referred to as the DF experiment), where they were tasked with analysing an evidence file obtained from the project Digital Corpora [27]. They were given a short scenario description and the same task: “What has happened, and what is Jean’s involvement in the reported incident?”. The participants solved the task individually, from their standard workstation, and could use hardware and software according to their preference. They were asked to spend four to five hours on the task, including analysis and report writing.

Due to the time limit for completing the DF experiment, reports that seemed or stated that they were incomplete regarding the content analysed in the current study were excluded (n=13). Only reports with a “conclusion” or “result” section (or variations of these) where the result of the analysis was summarized and/or conclusions were stated, were included (n=40) (See Table 1 for the demography of the DF examiners producing the reports).

It should be noted that the primary objective of the DF experiment was to establish whether DF examiners produce consistent information and whether contextual information biases the result. For exploring the issue of biasing contextual information, the participants were divided into four groups. The control group received the basic scenario, while the other three groups received variations of additional contextual information indicating guilt: a) Jean had confessed; b) Jean may have participated by helping an insider to leak the information; or innocence: c) Jean was framed in a phishing attack [11, Appendix 1]. The context had a significant effect on the observations of traces, but no biasing effect was found on interpretations or conclusions [11]. It is thus not considered likely that the context had a significant impact on the conclusion types examined in the current study. Regarding the content analysis, the current study centres on general content concerning the

Table 1
Demography of DF examiners in the current study.

Demography	n=40
Country	Norway: (34), India (1), The Netherlands (1), Denmark (1), Kenya (1), Canada (1), UK (1).
Gender	Female: 17.5%, Male: 82.5%.
Educational background	Civil: 50%, Police: 45%, Both 5%.
Educational level	PhD: 2.5%, MSc: 27.5%, BSc: 62.5%, Other 7.5%.
Organisational level	National/Fed./State: 40%, Local: 55%, Private/Independent: 5%.
Experience with DF within law enforcement	0-1 years: 27.5%, 2-4 years: 32.5%, 5-9 years: 32.5%, 10-15 years: 7.5%.

evidence credibility and is unrelated to the specific scenario information. It is therefore considered unlikely that the contextual information has impacted the content that was analysed here.

The conditioning information presented to the DF examiners indicated that their task was performed in an investigative phase. Jackson et al. [28] proposed three types of FS opinions: investigative opinions (expressed as categorical conclusions, explanations, or posterior probabilities of explanations), preliminary evaluative opinions, and fully evaluative opinions (expressed in probabilistic terms/LR). Due to the conditioning information provided, it was expected that the DF examiners would perceive the task as a part of the investigation phase and that their opinions would be either investigative opinions or preliminary evaluative opinions.

The task description was quite general, and due to the queries “what has happened” and “what is Jean’s involvement”, it was likely to prompt the participants to address source and activity level issues. The task description did not explicitly address offence level issues, but its general nature did not exclude opinions addressing such issues.

When studying DF reporting practices, it is relevant to compare with other FS disciplines, where e.g., Bayesian and frequentist methods have been used over a longer time than the relatively young DF discipline. Bali et al. [24,26] conducted a study of 500 reports sourced from forensic proficiency tests from eight different FS disciplines: Glass analysis, Paint analysis, Fibres analysis, Handwriting analysis, Questioned documents, Shoeprint impression, Toolmarks examination, and Firearms examination. The median number of reports from the included disciplines was 56, ranging from 36 to 121 reports [26], and the sample of 40 DF reports is thus considered useful for comparison. Bali et al. [24] describe that the reports were obtained from the 2016 sample, and according to the database (<https://www.cts-forensics.com/index-forensics-testing.php>), there were one or two different cases per discipline this year. The DF report sample originated from a single case scenario and would thus be suitable for comparison.

2.2. Analysis procedure

The analysis procedure for exploring DF reporting practices was threefold. First, the procedure for quantitative contents analysis of conclusion types and level of issues according to Hierarchy of Issues (HoI) [9,29] is elaborated. Second, the procedure for the quantitative content analysis of content concerning the credibility of the reported results is outlined. Eventually, the qualitative content analysis of certainty descriptors from the conclusions is described. Examples of coding categories are provided under the respective sections. Examples obtained from the current study were used when available and marked with participant number, e.g., “(P1)”. Examples that originally were in the Norwegian language were translated by the author of this paper.

2.2.1. Analysis of conclusion types

Bali et al. [24,26] performed a quantitative content analysis of conclusions and content with relevance for the credibility of the result. The coding from this study was applied and expanded with additional coding categories for the current study of DF reports.

The conclusions were coded in IBM SPSS Statistics (Release 26.0.0.0. 64 bit edition) according to six conclusions types (with dichotomous observed/not observed values if values are not specified): Categorical conclusions (values: only traditional, only elaborated, both), Likelihood ratios (LR) (values: numerical, verbal, both), Random match probabilities (RMP), Likelihood of observed similarity statements (LoS), Strength of support (SoS) statements, and Source probabilities (SP) (values: numerical, verbal, both). These conclusion types were based on the work of Thomson and colleagues [30,31] and are explained in short below. See

Table 2
Example of conclusion types.

Conclusion type	Example
Categorical conclusion	(traditional) “From the analysis, it is concluded that Jean was a victim of a phishing attack” (P54). (elaborated) “It seems like the firm m57.biz was exposed to CFO-fraud” (P19).
Random match probability (RMP) Likelihood ratio (LR)	“the chance of observing the evidence if it originated from an unrelated individual is 1 in 20” [24, p. 219]. “The strength of evidence is low in light of the hypothesis that messages never existed on the smartphone” [19, p. 4].
Source probability (SP)	(Numerical) “there is 90% chance that the suspect is the source”, (Verbal) “it is highly probable that the suspect is the source” [24, p. 219].
Likelihood of observed similarity (LoS)	“the likelihood of observing this amount of corresponding detail when the evidence did not come from the same source is low” [24, p. 219].
Strength of support (SoS)	“This may strengthen a hypothesis about Jean being negligent by having poor security and control, and thus have been a victim of phishing and fraud” (P65).

Table 2 for examples from the respective conclusion types. *Categorical conclusion (about source)*: The conclusion that two traces, for example, latent fingerprints, have the same source has traditionally been described as ‘identification’ or ‘individualization’, and the conclusion that they have a different source is called ‘exclusion’ [30]. For disciplines that use forensic expertise to address activity or even offence level issues, such as the DF discipline, the Categorical conclusion would also revolve around *what activity took place* (e.g., was the image *deleted*) or *what offence took place* (e.g., *which cybercrime* do the traces point towards). Traditional Categorical conclusions would be articulated as dichotomous match/no match or exclusion/cannot be excluded statements [9]. Bali et al. [24] also refer to elaborated Categorical conclusions, which include non-absolute and moderate statements such as “could have been the source”.

Random match probabilities: This conclusion type involves examiners following a two-step analysis, where they first compare the items, and if they find no distinguishing features, they would estimate and report how frequently indistinguishable items would be observed in a reference population, which would explain the probative value of this finding [31].

Likelihood ratio: According to Thompson, [31, p. 778] “LRs represent the expert’s view of the relative probability of the observed features under the alternative hypotheses about the source of the traces”. The probability assessment may rely on empirical data and statistical modelling, but may also be a subjective evaluation based on their training and experience. In some instances, the LR can partly rely on empirical data and partly on training and experience [31].

Source probability: This conclusion type is a statement about the probability that two items have a common source [31].

Likelihood of observed correspondence/similarity: This conclusion type entails making qualitative statements about the likelihood that the observed correspondence between the items could arise randomly if the items do not have a common source. The evaluation may rest in whole or in part on an examiner’s subjective judgment. This statement is similar to the verbal statement about relative likelihood. However, the difference is that it addresses only the likelihood of the observed pattern under the hypothesis of different sources, which entails that it speaks only to the denominator of the likelihood ratio. This approach could be misleading in cases where the numerator of the LR may be significantly less than one [31].

Strength of support for proposition: This is a translation of the LR into a

verbal statement about the strength of the evidence for supporting a particular hypothesis. The conclusion type aims at clarifying the meaning of the LR. It originates from people often mistakenly believing that LR is a statement about the probability of a particular hypothesis being true, instead of the strength of evidence for supporting the hypothesis [31].

2.2.2. Hierarchy of issues

The conclusions from the Bali et al. [24] study were focused on feature comparison procedures and would thus primarily address source level issues [See 9,29]. To gain more knowledge about which issues DF conclusions address, the conclusions were also analysed for whether they addressed issues at source/sub-source level, activity level, or offence level (values: only crime category, only guilt/intention, both). Since it sometimes was difficult to differentiate between the justification of the conclusion or a separate conclusion, all the narrative within the conclusion was coded according to the addressed issue level, as long as it was related to aspects concerning relevance or credibility of the information. This implied that several levels of issues could be observed within the narrative of a single conclusion.

2.2.3. Analysis of report content

A quantitative content analysis was performed based on thirteen categories of content. Eight categories were obtained from Bali et al. [24] (with dichotomous observed/not observed values if values are not specified): description of the analysis methods used (values: none, vague, specific), information about the method/tool reliability, information about the method/tool validity, limitations of the methods or conclusions, explication that the conclusion is an opinion, reasoning or justification for the conclusion (e.g., information about frequency or similarities), any alternative explanations for the results and additional explanation of jargon or scientific terminology.

Five additional content categories were applied due to the relevance for the DF discipline: Description of analysis tools used (values: none, inaccurate/vague, specific), task description (values: correct, inaccurate/vague, none), description of one’s qualification or competence, description of received contextual information (values: correct, inaccurate/vague, none), time zone explanation (values: none, only abbreviation/name, abbreviation/name and explanation of how it should be understood compared to local time). See Table 3 for examples of the content categories.

2.2.4. Analysis of certainty expressions

In addition to the analysis of conclusion types and report content, a qualitative analysis of the conclusions was performed to explore the use of expressions describing uncertainty. The analysis was aimed at exploring diversity rather than frequency. Due to the challenges described in the previous section, all the narrative within the conclusions was included in this analysis as long as it was related to issues concerning relevance or credibility of the information that was addressed. An example of what was excluded is recommendations of further investigation steps: “It may be necessary to interview Alison, to clarify what happened before the mail correspondence” (P34). All the phrases from the conclusions describing uncertainty within the above-mentioned criteria were extracted and sorted, and duplicate phrasings were excluded. The results of the qualitative analysis of certainty descriptors are presented in section 3.4.

3. Analysis

In this section, the results will be presented, analysed, and discussed. The analysis aimed to gain knowledge about what characterises a DF

Table 3
Example of coding categories.

Content type	Example
Alternative explanation for evidence	“The use of a USB stick during the relevant time period has been discovered. This should be further investigated” (P3).
Reliability of methods, scientist or tool	“this procedure is known to have an error rate of...” [24, p. 219].
Validity of analysis method or tool	“this analysis method has been empirically validated” [24, p. 219].
Potential limitations to the analysis or final opinion	“There are several limitations. Most of the fields in a message header may be altered, including the ‘return path’ and ‘from’ fields. It is thus not certain that these fields are correct” (P15).
Methods description	(vague) “The content analysis was performed in the program Axiom Examine” (P28). (specific) “I used the program X-Ways 19.8. Chose to search for known file header signatures, verified file types, unpack compressed files and retrieve intern metadata from files” (P22).
Explicit opinion statement	“The author’s interpretation of findings: (...)” (P49).
Reasoning or justification for the final opinion	“The user Jean@m57.biz has sent the spreadsheet. The one requesting the spreadsheet has pretended to be alison@m57.biz, but has tried to hide his real address simsong@xy.dreamhostps.com” (P33).
Term explanation	“Hash: Unique fingerprint for a digital file. If the hash value is the same, this entails that the content is exactly the same. Filenames are not included in the calculation of the hash value” (P28).
Task description	(correct) “The author of the report was tasked with analysing the evidence file from Jean’s computer and find out ‘what had happened, and what is Jean’s involvement in the reported incident’” (P13). (incomplete/wrong) “It is suspected that phishing lead to that confidential information was leaked. My task is to find something that can support this theory, and possibly find out how the information was spread” (P4).
Description of tool used	(vague) “The analysis was performed with the software Magnet AXIOM” (P41). (specific) “The units were examined with the analysis software’s X-Ways Forensics 19.8 and Axiom 3.1.0.14142” (P1).
Description of qualification	“(anonymised name), CCME, ECIH, CEH” (P60). “(anonymised name), PhD” (P58).
Complete context information	See [11, Appendix 1].
Time zone explanation	(only abbreviation) “20.07.08 01:22:45 UTC did Jean receive an email that appeared as being sent from alison@m57.biz” (P35). (abbreviation and explanation concerning local time) “Timestamps: The analysis was performed in UTC+2 which is Norwegian summertime. Timestamps referred in this report is in this time zone” (P9).

opinion, and the study of FS opinions was used as a reference point. For clarity, the research questions are repeated here: First, what characterises the opinions used in DF reports in terms of opinion type, uncertainty expressions, addressed issues, and included content relevant to the credibility assessment of the reported results, and second, to what degree does the DF reporting practice concerning applied opinion types and included content deviate from the other FS disciplines.

First, the result and analysis from the quantitative content analysis of conclusion types and level of issue are presented, followed by the quantitative content analysis of report content and the qualitative content analysis of certainty expressions. Eventually, a summary discussion is offered, where also limitations and future directions are commented on.

Table 4
Proportions (%) of conclusion types from DF reports and FS disciplines, obtained from [24,26].

Conclusion type	Digital forensicsn=40	TotalFSn=500	Fibres	Fire-arms	Glass	Hand-writing	Paint	Docu-ments	Shoe-print	Tool-marks
Categorical	50	94.4								
<i>Traditional</i>	20	73.8	39.5	95.9	18.2	55.8	35.9	93.2	67.2	100.0
<i>Elaborated</i>	27.5	9.8	32.6	0	36.4	13.5	41.0	0	0	0
<i>Both</i>	2.5	10.8	27.9	1.7	42.4	17.3	7.7	0	21.9	0
RMP	0	0	0	0	0	0	0	0	0	0
LR	0	0.4	0	0	6.1	0	0	0	0	0
LOS	0	0.8	0	0.8	3.0	3.8	0	0	0	0
SP	0	0	0	0	0	0	0	0	0	0
SoS	50	4.4	2.3	0.8	12.1	11.5	7.7	6.8	4.7	0
Mean	12.5	12.5	12.8	12.4	14.8	12.7	11.5	12.5	11.7	12.5
SD	18.6	25.2	17.3	33.7	16.5	18.7	17.0	32.7	23.7	25.2

3.1. Conclusion types

Since the DF experiment scenario suggested that they were in an early phase of the investigation, it was expected that opinions would be articulated as either investigative opinions or preliminary evaluative opinions.

The observed conclusions types in DF reports were Categorical conclusion and Strength of support (SoS), with an even distribution between the two conclusion types. Three of the disciplines in the Bali et al. [24] study also used the Likelihood of observed similarity (LoS) conclusion type, and one discipline used the Likelihood ratio (LR) conclusion type, see Table 4.

A significance level of 0.01 was used in the statistical tests. Since there was no related prior literature on DF vs FS opinions and report content, Cohen’s benchmarks (1988) was used as reference for the effect size (phi: small/weak effect 0.1, medium/moderate effect 0.3, large/strong effect 0.5, Cramer’s V: small/weak effect 0.2, medium/moderate effect 0.5, large/strong effect 0.8).

The Categorical conclusion type was observed more often in the FS conclusions (94.4 %) compared to the DF conclusions (50%). More specifically, the *traditional* type was more often observed in FS conclusions (73.8%) compared to the DF conclusions (20 %), while the *elaborated* type was more often observed in the DF conclusions (27.5 %), compared to the FS conclusions (9.8 %). The combination of *traditional* and *elaborated* types was observed in a quite low proportion of the DF and FS conclusions.

A Pearson Chi-Square showed a significant association with moderate effect between discipline type (DF/FS total) and the *Categorical conclusions* ($X^2 = 110.443$, $df = 3$, $p = .000$, Cramer’s $V = .452$). When comparing the DF reports and the reports from the individual FS disciplines, the chi-squared test showed significant associations with moderate or strong effects between discipline type (DF/FS) and the *Categorical conclusions* (see Appendix 1). The results indicate that the reports from the included FS disciplines are associated with a higher likelihood of including *Categorical conclusions*, particularly the *traditional* type compared to DF reports. The effect of the association varied from moderate to strong.

According to Bali et al. [24], the Categorical conclusion is the conclusion type that has received the most criticism in recent years. Traditional Categorical conclusions have been criticized since they imply absolute certainty or a greater level of certainty than is warranted [30,32–34]. According to Thompson et al. [30], Categorical conclusions have, on the other hand, been defended by arguing that strong statements are necessary to convey the high probative value of an expert’s conclusion and that legal decision-makers would give too little weight to conclusions that were articulated otherwise.

The elaborated Categorical conclusions have many similarities with

what is referred to as explanations in the Case Assessment and Interpretation (CAI) framework [9,35], since they were expressed with no evaluation of the strength of the proposed explanation, with descriptors such as “this may be...”, “could have...” or “there are traces of...” [36]. Such opinions may be helpful as investigative advice aimed at guiding the further search for evidence in the investigation. However, the usefulness of such opinions is debated, and it is claimed that they may be impractical or even whimsical suggestions, since investigating multiple and poorly founded explanations introduced by forensic examiners may require resources from the investigation team that otherwise could have been used more effectively [36].

The use of Bayesian methods such as LR in forensic evaluation has been recommended in guidelines for the DF discipline [37] and also demonstrated in several research papers [19–21]. Despite this, none of the DF reports were articulated as preliminary evaluative opinions with LR.

The *Strength of support* (SoS) conclusion type is used to describe to what degree the result supports a proposition or hypothesis. SoS was observed more often in the DF conclusions (50%) compared to the FS conclusions (4.4 %). A Pearson Chi-Square Test showed a significant association with moderate towards strong effect between discipline type (DF/FS total) and the SoS conclusion type ($X^2 = 107.368$, $df = 1$, $p = .000$, $\phi = .446$), and significant associations with moderate or strong effects between DF discipline and the individual FS disciplines for SoS conclusion type (see Appendix 1). The results indicate that the DF reports are associated with a higher likelihood of using SoS conclusions than reports from the other FS disciplines included here and that the effect of the association varied from moderate to strong.

It should be noted that all SoS conclusions in the DF reports referred to a single hypothesis/explanation and that none referred to a set of propositions or multiple competing hypotheses. Descriptors such as “indicates...”, “the findings suggest...”, or “no indications of...” were used and could thus be the result of an implicit evaluation of the probability in light of a single hypothesis. The conclusion type has similarities with what is referred to as posterior probabilities in the CAI framework [9,36]. Deciding on a posterior probability involves evaluating the evidence in light of the hypothesis while considering the prior probability. A prior probability is the probability of the evidence in light of the hypothesis before considering the examination result, such as underlying frequencies of occurrences of an event. The strength of support conclusions or posterior probability opinions should thus exercise transparency about the prior probabilities. Otherwise, it would be unclear how the examiner arrived at the posterior probability.

An example of failing to convey priors can be found in the DF experiment reports: Before the analysis, the DF examiners were informed that the boss had asked the suspect to create a spreadsheet with information about the employees, which entails that the chance of

Table 5
Proportions (%) of opinion type in DF reports.

Opinion type	DF reports (n=40)	Sourcelevel issue	Activity level issue	Offence level issue
Categorical conclusion	50			
- traditional	20	17.5	17.5	10
- elaborated	27.5	27.5	25	20
- both	2.5	2.5	2.5	0
LR	0	0	0	0
RMP	0	0	0	0
LoS	0	0	0	0
SP	0	0	0	0
SoS	50	47.5	45	32.5
Total		95	90	62.5

finding the information on the computer is high regardless of whether the suspect was involved in leaking the spreadsheet to someone outside the firm or not. Some of the DF examiners who concluded that the suspect had committed the information leakage justified the conclusion with the discovery of the spreadsheet without mentioning that the suspect had a legitimate reason for having it on her computer. Such a conclusion could mislead legal decision-makers to believe this was incupating evidence when it, in fact, was neutral.

There are some risks with the SoS approach used in the DF reports that should be highlighted. First, when the evidence is presented in relation to a single hypothesis, the conclusion may be perceived as the probability that the hypothesis is true, known as the prosecutor’s fallacy [38] or the fallacy of transposed conditions [39]. Second, by presenting evidence in relation to a single hypothesis, the legal decision-maker may be misled to believe that the evidence has probative value, while an evaluation in the light of several competing hypotheses would show that it was neutral (not suitable to discriminate between propositions or a set of hypotheses) [40], as shown in the example above. If the DF examiners presented the evidence relative to a proposition or a set of competing hypotheses, the chance of such misconceptions would thus be reduced.

3.2. Hierarchy of Issues (HoI)

The DF conclusions were analysed for which level within the HoI [9,29] they addressed. This content was not addressed in the study of FS disciplines [24] and comparison was therefore not performed.

The digital trace would be central for issues at all HoI levels. A trace is defined by Pollitt et al. as “any modification, subsequently observable, resulting from an event” [18, p. 1]. At the source level, issues would centre on which source or identity-related information a trace or an event originated from, for example, trace/event– user account (sub-source), or trace/event – user account (sub-source) – person (source). An example from the current study was “If the user account ‘Jean’ on the analysed system is Jean Jones, the findings mentioned here indicate that Jean may have sent the file ‘m57biz.xls’ per email to tuckergorge@gmail.com or alison@m57.biz on 2008-07-20 03:28:00 CEST” (P42). Although the example was concerned with multiple issues, the source level issue here would be about the connection between the file m57.biz (trace) – jean user account (sub-source) – Jean (source).

Activity level issues would be concerned with linking activities or events to a trace, e.g., creating a file, deleting a log, downloading an image, sending an email, or executing a program. An example from the current study is “The Excel-file m57biz.xls was last changed by Jean Jones” (P2).

Offence level issues would be about the legal offence concepts that constitute a crime or concerned with the subjective side of the crime, related to motivation, intentions, or guilt. An example of an offence level issue involving offence concept was “From the analysis it is concluded

Table 6
Proportions (%) of content type in DF reports/conclusions.

Content type	DF reports(n=40)
Term explanation	15
Task description	75
Specific - correct	52.5
Inaccurate/vague	22.5
Description of received context	40
Specific - correct	12.5
Inaccurate/vague	27.5
Description of tools used	70
Specific	57.5
Inaccurate/vague	12.5
Qualification or competence	5
Time zone	70
Time zone abbreviation and explanation	15
Time zone only name/abbreviation	55

that Jean is a victim of a phishing attack” (P54). An example involving criminal guilt was “Jean probably has in good faith sent the attachment which is posted on the internet to the one she thought was Alison Smith” (P16).

The results are presented in Table 5 and show that all three HoI levels were addressed in the DF conclusions, regardless of conclusion type. 95% of the DF conclusions addressed source level issues, 90% addressed activity level issues, and 62.5% addressed offence level issues. The conclusions that addressed the offence level were further analysed for whether they addressed the crime concept, intentions/guilt, or combinations of these. The crime concept was addressed in 20% of the conclusions, intentions/guilt was addressed in 15% of the conclusions, and 27.5% addressed both crime concept and intentions/guilt.

Within FS in general, the offence level is considered outside the domain of scientists [29, p. 233]. However, the current study suggests that the DF discipline has a different practice concerning this issue. It is also contested whether a scientist should address the activity level issues due to the need to obtain and consider case-specific information which is outside the scientist’s expertise [41]. While this may be correct for other FS disciplines, this is not necessarily transferable to the DF discipline. As demonstrated by Henseler and de Poot [42], the digital information in DF analyses often contain information that can be used to reconstruct and prove criminal offences, not only at source level, but also at activity level. The high proportion of activity and offence level issues in DF conclusions observed in the current study suggests that DF may deviate from other FS disciplines in this respect.

Since the reports included in this study originated from an experimental setting, it is relevant to explore in future studies to what degree the results are representative for DF reports in real casework. It is also relevant to ask whether there are sufficient knowledge and awareness among DF examiners about which level of the HoI they address and the uncertainty that the respective level introduces that must be taken into account. A broader knowledge base may facilitate a normative discussion about how, when, and by whom (which expertise) digital evidence at the respective HoI levels should be examined and evaluated. Such knowledge is of great value for the development of procedures and standards aimed at supporting safe analysis and reporting practices in the DF discipline, which may prevent errors of justice due to misinterpretation of relevance and probative value of digital evidence by legal decision-makers.

3.3. Content analysis

The eight different content types from the study of FS disciplines [24,26] were included in the content analysis, see Table 7. They were supplemented with five additional content types, see Table 6. The

Table 7
Proportions (%) of content type in reports/conclusions in the DF discipline and FS disciplines obtained from [24,26].

Content type	Digital forensicsn=40	TotalFSn=507	Fibres	Fire-arms	Glass	Hand-writing	Paint	Docu-ments	Shoe-print	Tool-marks
Alternatives	32.5	9.6	18.6	13.2	21.2	0	30.8	6.8	0	0
Reliability	0	0.2	0	0	0	0	2.6	0	0	0
Validity	0	0.2	0	0	0	0	2.6	0	0	0
Limitations	60	5.2	7	4.9	6.1	17.3	5.1	6.8	0	0
Methods vague	55	11.2	7	14.9	0	3.8	10.3	20.3	12.5	10.1
Methods specific	17.5	18.4	18.6	23.1	24.2	3.8	15.4	20.3	4.7	28.1
Opinion statement	2.5	9.8	4.7	6.6	12.1	17.3	10.3	20.3	7.8	5.6
Reasoning	40	46.2	67.4	43.8	60.6	30.8	46.2	54.2	42.2	40.4
Term explication	15	0.8	2.3	0.8	0	1.9	0	0	1.6	0
Mean	24.7	11.3	14	11.9	13.8	8.3	13.7	14.3	7.6	9.4
SD	23.2	14.4	21.2	14.3	19.9	10.9	15.3	17.4	13.7	14.9

additional content types were selected due to their importance to understanding the relevance and meaning of the digital evidence, and for the credibility assessment of the result.

Some of the content types were observed in a high proportion of the reports. However, the content analysis of the DF reports suggests that important information for assessing the credibility of the result was often absent or inadequately described.

A description of tool(s) used was found in 70% of the DF reports; however, only 57.5% were specific and included version numbers. 70% included the time zone abbreviation, although only 15% explained how it should be understood in contrast to local time.

Due to the experimental setting, the ground truth was known for some content categories, such as the task descriptions and received context, which enabled a validity assessment. A correct task description is crucial for assessing whether the DF examiner chose adequate tools and methods for the DF examination and whether the conclusion provides an adequate answer to what was requested. Only about half of the DF examiners (52.5 %) provided an accurate task description, while 47.5 % did not provide one or provided a vague/inaccurate task description.

The DF experiment showed that the DF examiners were biased by the contextual information they received and that it impacted how many traces they discovered and deemed relevant to include in their analysis reports [11]. To minimize cognitive bias, it is therefore essential that the DF examiner only receives task-relevant information. Insight into what the DF examiner knew when the examination was done is vital for evaluating whether the results may have been biased in the wrong direction [11]. Obtaining oversight over what information the DF examiner received can be challenging in a real-life investigative setting due to the widespread use of dialogue or discussion with the investigation team to settle the task and scope of the DF examination [11]. In the DF experiment, the context was provided in a written form and would be straightforward for the DF examiners to present accurately in the report. Hence, it is of concern that only 40 % described the context they received, and that only 12.5% of these descriptions were correct and accurate.

A DF report is always written for someone else and usually for someone with low-level technical knowledge. Explanation of the terminology used in the report is thus crucial for understanding what is presented in the report. It is therefore alarming that only 15 % of the DF reports included such explanations.

The description of qualification or competence was only included in only 5 % of the DF reports. Such information may indicate whether the DF examiner holds sufficient expertise for the methods and tools used, and for interpreting the results. Due to the complex and rapidly changing nature of technology compared to the more stable scientific underpinning in other FS disciplines, this information is both relevant

and necessary for assessing the credibility of the result. The very low proportion of DF examiners conveying this information is a cause for concern.

The rest of the content types were suitable for comparison between the DF and FS disciplines, and are presented in Table 7.

A Pearson Chi-Square Test showed a significant association with moderate towards strong effect between discipline type (DF/FS total) and description of limitations ($X^2 = 134.409$, $df = 1$, $p = .000$, $\phi = .496$) and descriptions of methods (vaguely) ($X^2 = 50.449$, $df = 1$, $p = .000$, $\phi = .324$), and a significant association with moderate towards weak effect between discipline type and explicated terms ($X^2 = 41.719$, $df = 1$, $p = .000$, $\phi = .276$) and alternative explanations on the evidence ($X^2 = 20.486$, $df = 1$, $p = .000$, $\phi = .194$). The chi-squared tests between discipline and other categories showed non-significant results – see Appendix 1. The results indicate a higher likelihood of finding descriptions of limitations and methods (vague), explicated terms and alternative explanations of evidence in DF reports compared to reports from FS disciplines included here.

A detailed analysis of comparison of DF and FS disciplines on the individual content types was not performed. However, a more general comparison of mean proportion of content suggests that the reports from the DF discipline ($M = 24.7$, $SD = 23.2$) on average had a higher proportion of the compared content than the included FS disciplines ($M = 7.6 - 14.3$, $SD 10.9 - 21.2$).

There were substantial deficiencies concerning method descriptions, statements about reliability/validity of methods or tools, and explicating that the conclusion was an opinion. These challenges do not seem exclusive to the DF discipline since similar tendencies were also found within the FS disciplines [24,26].

A detailed method description is important for verification and peer review purposes but was provided in only 17.5% of the DF reports. Even more alarming, none of the DF examiners commented on the reliability or validity of methods or tools in their reports. Due to the rapid technology development, no tool can be trusted to perform perfectly over time and should be tested regularly. Information about whether there are known errors associated with the tool is vital for assessing the credibility of the result. The missing content concerning the reliability and validity of tools used in the analysis is thus of concern.

Only 2.5% of the DF reports explicated that the conclusion was an opinion, which is problematic, since digital evidence may easily be associated with being objective and credible by default of those with low-level technical knowledge. There is a risk that opinion statements may be perceived as factual statements by legal decision-makers and that the uncertainties associated with the opinion evidence are not acknowledged.

3.4. Certainty expressions

The qualitative content analysis showed that the following expressions were used to convey uncertainty in the DF report conclusions, and related types of expressions are presented together in bullet points:

- Cannot see an overwhelming probability of...
- It is not possible to conclude with...
- Cannot exclude...
- No indications of...
- It can be speculated...
- It is uncertain whether...There is uncertainty with...
- Probably...It is probable...It is very probable...It is highly probable...
- Possible...It is still possible...It is possible...Possibly...
- Apparently...It is apparent...
- It appears as...It may appear as...Appears to be...Appears as if...
- Conceivable...It is conceivable
- May be...Might have...Could have...May be related to...
- There are traces of...There are signs of...
- Reason to believe...Great reason to believe...
- Indicate...This may indicate...The finding suggest
- Which largely corresponds with...
- It is clearly accurate...
- It is provable...

The analysis of DF reports shows that uncertainty was articulated with a wide variety of expressions. None of the DF examiners described how to interpret these expressions. Some expressions may have similarities with descriptors used in established frameworks for evaluative reporting [43], for example, “is very probable” and “highly probable”. However, none of the reports referred to an established framework of expressions, which would provide more clarity in understanding these expressions’ probative value.

A large proportion of the expressions would fall into the category of “indeterminate opinions”, which are characterised as vague, ambiguous, and difficult to understand [9, p. 127]. For example, the expression “cannot exclude” may be used when the examiner considers the hypothesis an inadequate explanation for the findings, but cannot entirely refute it as a possibility. The problem is, without providing a qualification, the legal decision-maker may get the impression that the suggested explanation is a good one [36]. Indeterminate descriptors should be avoided due to the risk of misleading the legal decision-maker to misinterpret the probative value of the evidence, which in turn may cause erroneous legal decisions.

3.5. Summary discussion, limitations and future directions

The primary aim of this study was to gain more knowledge about what characterises the opinions used in DF reports in terms of opinion type, uncertainty descriptors, addressed issues, and inclusion of content with relevance to the credibility assessment of the reported results. The study suggests that DF examiners mainly tend to use two opinion types in their reports: Categorical conclusion and SoS. The conclusions addressed all levels of HoI, with a surprisingly large proportion of issues at offence level, which in other FS disciplines is regarded as an issue outside the scope of the forensic scientist. The main finding in the report content analysis is that there seem to be several deficiencies in the reporting practices, where vital information is missing, presented vaguely, or even erroneously. The great variety of certainty expressions in use, combined with no explication of meaning or reference to a standard framework, indicate that standards such as the Certainty scale

[19,44–45] or the Digital Evidence Certainty Descriptors [46] are not known or implemented sufficiently.

The secondary aim of this study was to determine to what degree opinions and content presented in DF reports deviate from the practices of other FS disciplines. The analysis shows that even if a few FS reports included other conclusion types (LoS and LR), the vast majority of conclusions from both the FS and DF reports were Categorical conclusions and SoS. The analysis indicates a higher likelihood of observing Categorical conclusion type and particularly the traditional type in FS reports than in DF reports, while a higher likelihood of observing SoS conclusion types in DF reports compared to FS reports. The associations between discipline type and these conclusion types were significant, with moderate or strong effect sizes.

When comparing report content, the study suggests a higher likelihood of observing descriptions of limitations and methods (vague) in DF reports compared to reports from the FS disciplines included here. The association between disciplines and content types was significant with moderate towards strong effect size. There was also a higher likelihood of observing term explanations and alternative explanations of evidence in DF reports compared to FS reports, and although the association between discipline and content type was significant, the effect size was moderate towards weak. The analysis showed that a higher proportion of FS reports included reasoning or justification of the conclusion and explicated that the conclusion was an opinion compared to reports from the DF discipline, however, the statistical tests showed that these were weak and non-significant associations.

When considering the lack of balanced conclusions among the DF examiners (none were presented in relation to multiple hypotheses), the deficiencies found in the content analysis, and the insufficient use of certainty descriptions in the DF reports, there seems to be a high risk of erroneous or misleading results presented in DF reports. The current study indicates a substantial and urgent need to increase the quality of DF reports and conclusions.

3.5.1. Validity and limitations

The sample and material are considered adequate and representative of current DF practice. The participants were all DF practitioners, with a variation in gender, organisational level, educational background, competence, and experience with DF work, as shown in Table 1. The DF experiment had high ecological validity due to an experimental design that mimicked regular DF casework: They received some background information about the case, an evidence file, and a task description. They processed and analysed the evidence file with tools and methods they would typically use and were encouraged to use their standard structure or template for the report.

There are, however, some possible limitations that should be discussed. First, similar to the study of FS reports [24], the DF reports were not collected from actual casework. Since the DF reports were obtained from an experimental setting, the DF examiners were aware that a researcher would examine their reports. This circumstance may have led to a higher focus on quality compared to how they would approach an average case. On the other hand, they knew that the reports would not undergo scrutiny in a court hearing, which may have impacted the quality in the opposite direction. Based on these conditions, it is not possible to determine which direction the experimental setting may have skewed the results. It would thus be very useful to compare with reports collected from actual DF casework in future studies.

Second, the scenario information indicated that they were at the investigation stage, as opposed to a request for an evaluative opinion from the court. Due to this setting, the findings related to conclusion types may have been skewed. If the DF reports were from the court stage, more opinions of LR type might have been observed. To explore whether

they were biased by contextual information, the participants (except the control group) were given different contexts in addition to the scenario information [11]. Bias was observed at the observation level, but the study did not suggest bias on the interpretation or conclusion levels. However, absence of evidence does not entail evidence of absence, and biased interpretations and conclusions can therefore not be excluded.

Third, the task description “*What has happened, and what was Jean’s involvement in the reported incident?*”, was open and general regarding the HOI level(s) it addressed. The task description would not limit the practitioner to provide results addressing a particular level and would, as explained in section 2.1, probably prompt conclusions at source and activity level. The task description did not explicitly address the offence level. However, the phrase “what has happened” could be associated with determining what crime concept had occurred, which would be an offence level issue. The task description did not ask the participants to evaluate guilt or innocence, and the offence level issues related to intentions/guilt issues observed in the reports can therefore not be directly attributed to the task description.

Forth, the limited timeframe for the DF examiners to conduct the analysis and write a report may have impacted the result. To mitigate this limitation, only reports that contained a conclusion were included, and reports that seemed or stated that they were incomplete were excluded. Despite the limited timeframe, it would be expected that most of the content categories examined here (e.g., term explanation, explanation that the conclusion is an opinion, description of reliability or validity of methods/tools and qualification/competence) should be a part of a standard procedure, and available in the template or checklist for DF report writing. If so, such content would not need to be articulated every time but instead adjusted to fit well with the particular report. Therefore, the DF report content deficiencies may be explained by an insufficient implementation of best practices rather than time limitations.

3.5.2. Future directions

In terms of future directions, the inclusion criteria in the Bali et al. [24] study was feature comparison disciplines, which entails that they may have excluded several disciplines that would have been relevant for comparison with the DF discipline. The current study suggests that the DF discipline may be characterised as an exploratory or investigative discipline, not only tasked with classification or identification processes for issues at source level, but also with reconstruction processes for issues at activity and offence level. In future studies, it would be relevant to compare with more similar disciplines, such as bloodstain pattern analysis, crime scene analysis, and forensic pathology.

More research on DF practices is needed to facilitate clear, accurate, and comprehensible communication of DF opinion evidence to legal decision-makers and prevent potentially misleading DF evidence. A future aim may be to provide a normative framework for presenting opinion evidence in the DF discipline. To get there, it is necessary to obtain more knowledge about the DF examiners’ *practice* when they interpret, evaluate and present opinions about digital evidence, and *why* they do what they do concerning factors such as knowledge and motivation.

4. Conclusion

To the best of the author’s knowledge, this is the first study to explore DF reports and opinion evidence and is therefore an important contribution to the sparse but growing body of research about DF practices. Some limitations, such as the experimental setting, the indications of an investigation stage (as opposed to an evaluation stage), the task

description, and the time limitation, were identified and discussed.

The key findings from the study are that the most frequently observed conclusion types in the DF reports are Categorical conclusions and SoS. These conclusion types were the most commonly used by the FS disciplines included in [24,26]. However, there were a few instances of other conclusion types (LR, LoS) used by some FS disciplines. A comparison of the disciplines suggests that the FS disciplines included in [24,26] used Categorical conclusions and particularly the traditional type significantly more often than the DF discipline and that the DF discipline used SoS conclusion type significantly more often than the FS disciplines. The traditional Categorical conclusion type is contested since it concludes without conveying any uncertainty. The SoS conclusion type is also problematic since it may be one-sided and misleading if it only relates to one hypothesis, which was the situation in the DF conclusions. Despite the call to be cautious when choosing HoI level for the opinions, the DF conclusions addressed all three levels, namely the source, activity, and offence levels.

The content analysis suggests several challenges with DF reporting practices and that crucial information for assessing the credibility of the result is often absent or inadequately presented. Of particular concern is that all DF reports lacked descriptions of validity or reliability of methods, that the majority did not describe the contextual information they had received, that about half did not provide an accurate and correct task description, and that almost a third had no description of the methods or tools used. Very few explicated that the conclusion was an opinion, described their qualification/competence, or explained the terminology used in the report. The comparison of the DF and FS disciplines showed that they share several of the identified challenges, such as terminology explanations and describing validity and reliability of methods/tools.

The analysis of certainty expressions revealed a high variation in how uncertainty is conveyed in DF reports. Due to the missing elaboration of how the certainty expressions are to be understood or the lack of reference to an established framework, there is a risk of misleading the legal decision-makers about the probative value of the evidence.

To summarise, the study provides valuable and relevant insights into DF reporting practice. The choice of conclusion types, the deficiencies in the reporting practices regarding content and conveying uncertainty should be of concern. The results indicate that more research on evaluation and reporting practices in the DF discipline is necessary and justify a need for enhanced focus on developing safe and high-quality reporting practices. The findings also justify a need for enhanced focus on quality control and peer review within the DF discipline.

Funding source

This paper was funded by a grant from the Norwegian Police University College awarded the author for pursuing a PhD.

CRediT authorship contribution statement

Nina Sunde: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing.

Acknowledgements

I would like to thank Professor Helene O. I. Gundhus, Professor Johanne Y. Dahl, Associate Professor Fergus Toolan, and Professor Torbjørn Skardhamar for insights and constructive criticism of former versions of the paper. I also thank the editor and the anonymous reviewers for their helpful suggestions and inputs.

Appendix 1. - Chi-squared statistics and effect sizes for associations between disciplines and conclusion types / content types

Conclusion types	Degrees of freedom	Pearson Chi-Square	Asymptotic Significance (2-sided)	Effect size (Phi) ^a / (Cramer's V) ^b
DF total vs FS total (n=540)				
Categorical conclusions	3	110.443	.000	.452 ^b
SoS	1	107.368	.000	.446 ^a
DF vs FS - FIBRES ANALYSIS (n=84)				
Categorical conclusions	3	29.976	.000	.597 ^b
SoS	1	25.455	.000	.550 ^a
DF vs FS - FIREARMS EXAMINATION (n=160)				
Categorical conclusions	3	106.833	.000	.817 ^b
SoS	1	63.602	.000	.630 ^a
DF vs FS - GLASS ANALYSIS (n=79)				
Categorical conclusions	3	17.845	.000	.475 ^b
SoS	1	14.747	.000	.432 ^a
DF vs FS - HANDWRITING EXAMINATION (n=93)				
Categorical conclusions	3	22.983	.000	.497 ^b
SoS	1	16.933	.000	.427 ^a
DF vs FS - PAINT ANALYSIS (n=76)				
Categorical conclusions	3	15.961	.001	.458 ^b
SoS	1	15.587	.000	.453 ^a
DF vs FS - QUESTIONED DOCUMENTS (n=99)				
Categorical conclusions	3	56.152	.000	.753 ^b
SoS	1	24.247	.000	.495 ^a
DF vs FS - SHOEPRINT IMPRESSION (n=100)				
Categorical conclusions	3	57.137	.000	.756 ^b
SoS	1	27.442	.000	.524 ^a
DF vs FS - TOOLMARKS EXAMINATION (n=129)				
Categorical conclusions	3	94.689	.000	.857 ^b
SoS	1	52.665	.000	.639 ^a

*The number of cases (n) concerning FS conclusion types were obtained from the corrected Table 3 in [26].

Content types	Degrees of freedom	Pearson Chi-Square	Asymptotic Significance (2-sided) ^a / Fisher's Exact Test (2 sided) ^b	Effect size (Phi)
DF total vs FS total (n=547)				
Alternative explanations	1	20.486	.000 ^b	.194
Reliability	1	0.079	1.000 ^b	.012
Validity	1	0.079	1.000 ^b	.012
Limitations	1	134.409	.000 ^b	.496
Methodology vague	1	50.449	.000 ^a	.324
Methodology specific	1	0.018	.894 ^a	.006
Opinion statement	1	2.990	.107 ^b	.074
Reasoning	1	0.496	.481 ^a	.030
Term explanation	1	41.719	.000 ^b	.276

*The number of cases (n) concerning FS content types were obtained from the corrected Table 4 in [26].

References

[1] A. Giddens. *The consequences of modernity*, Stanford University Press, Stanford, 1997.

[2] B. Carrier, Open source digital forensic tools: The legal argument. https://dl.packetstormsecurity.net/papers/IDS/atstake_opensource_forensics.pdf, 2002 (accessed 4 July 2021).

[3] J.T. McDonald, Y.C. Kim, A. Yasinsac, Software issues in digital forensics, *ACM SIGOPS Operating Systems Review* 42 (3) (2008) 29–40, <https://doi.org/10.1145/1368506.1368512>.

[4] J.S. Atkinson, *Proof is not binary: The pace and complexity of computer systems and the challenges digital evidence poses to the legal system*, *Birkbeck Law Review* 2 (2014) 245–262.

[5] M. Rogers, *Technology and digital forensics*, in: M.R. McGuire, T.J. Holt (Eds.), *The Routledge handbook of technology, crime and justice*, Routledge, Oxon, 2017, pp. 406–416.

[6] W. A. Bhat, A. AlZahrani, M.A Wani, Can computer forensic tools be trusted in digital investigations?, *Science & Justice*, 61(2) (2021) 198-203. <https://doi.org/10.1016/j.scijus.2020.10.002>.

[7] N. Sunde, Non-technical sources of errors when handling digital evidence within a criminal investigation, Master's thesis, The Norwegian University of Science and Technology, 2017.

[8] N. Sunde, I.E. Dror, Cognitive and human factors in digital forensics: Problems, challenges, and the way forward, *Digital Investigation*, 29, (2019) 101-108. <https://doi.org/10.1016/j.diin.2019.03.011>.

[9] G. Jackson, C. Aitken, P. Roberts, Case assessment and interpretation of expert evidence. Guidance for judges, lawyers, forensic scientists and expert witnesses. Practitioner guide No 4. Royal Statistical Society. <http://www.rss.org.uk/Images/PDF/influencing-change/rss-case-assessment-interpretation-expert-evidence.pdf>, 2015 (accessed 4 July 2021).

[10] N.M. Smit, R.M. Morgan, D.A. Lagnado, A systematic analysis of misleading evidence in unsafe rulings in England and Wales, *Science & Justice*, 58(2) (2018) 128-137. <https://doi.org/10.1016/j.scijus.2017.09.005>.

[11] N. Sunde, I.E. Dror, A Hierarchy of Expert Performance (HEP) applied to digital forensics: Reliability and biasability in digital forensic decision making, *Forensic Science International: Digital Investigation* 37 (2021) 301175, <https://doi.org/10.1016/j.fsidi.2021.301175>.

[12] P. Sommer, Accrediting digital forensics: what are the choices?, *Digital Investigation*, 25 (2018) 116-120. <https://doi.org/10.1016/j.diin.2018.04.004>.

[13] B. Cusack, Extracting benefits from standardization of digital forensic practices, *Policing: A Journal of Policy and Practice* 15 (1) (2021) 59–67, <https://doi.org/10.1093/police/paz064>.

[14] J.H. Jahren, Is the quality assurance in digital forensic work in the Norwegian police adequate?, Master's Thesis, Norwegian University of Science and Technology, 2020.

[15] H. Page, G. Horsman, A. Sarna, J. Foster, A review of quality procedures in the UK forensic sciences: What can the field of digital forensics learn?, *Science & Justice*, 59(1) (2019) 83-92, <https://doi.org/10.1016/j.scijus.2018.09.006>.

[16] G. Horsman, N. Sunde, Part 1: The need for peer review in digital forensics, *Forensic Science International: Digital Investigation* 35 (2020) 301062, <https://doi.org/10.1016/j.fsidi.2020.301062>.

[17] N. Georgiou, R.M. Morgan, J.C. French, Conceptualising, evaluating and communicating uncertainty in forensic science: Identifying commonly used tools through an interdisciplinary configurative review, *Science & Justice*, 60(4) (2020) 313-336. <https://doi.org/10.1016/j.scijus.2020.04.002>.

[18] M. Pollitt, E. Casey, D.O. Jaquet-Chiffelle, P. Gladyshev, 2018. A framework for harmonizing forensic science practices and digital/multimedia evidence, OSAC Technical Series 0002, OSAC/NIST. https://www.nist.gov/system/files/documents/2018/01/10/osac_ts_0002.pdf (accessed 4 July 2021).

- [19] E. Casey, Standardization of forming and expressing preliminary evaluative opinions on digital evidence, *Forensic Science International: Digital Investigation* 32 (2020) 200888, <https://doi.org/10.1016/j.fsidi.2019.200888>.
- [20] E. Casey, D.O. Jaquet-Chiffelle, H. Spichiger, E. Rysler, T. Souvignet, Structuring the evaluation of location-related mobile device evidence, *Forensic Science International: Digital Investigation* 32 (2020), 300928, <https://doi.org/10.1016/j.fsidi.2020.300928>.
- [21] M. Tart, Opinion evidence in cell site analysis, *Science & Justice*, 60(4) (2020) 363–374. <https://doi.org/10.1016/j.scijus.2020.02.002>.
- [22] S. Doyle. *Quality management in forensic science*, Academic press, London, 2019.
- [23] C.E. Berger, J. Buckleton, C. Champod, I.W. Evett, G. Jackson, Response to Faigman, et al., *Science & Justice* 51 (4) (2011) 215, <https://doi.org/10.1016/j.scijus.2011.09.006>.
- [24] A.S. Bali, G. Edmond, K.N. Ballantyne, R.I. Kemp, K.A. Martire, Communicating forensic science opinion: An examination of expert reporting practices, *Science & Justice*, 60(3) (2020) 216–224. <https://doi.org/10.1016/j.scijus.2019.12.005>.
- [25] C.J. Lawless, R. Williams, Helping with inquiries or helping with profits? The trials and tribulations of a technology of forensic reasoning, *Social Studies of Science*, 40 (5) (2010) 731–755. <https://doi.org/10.1177/0306312710378787>.
- [26] A.S. Bali, G. Edmond, K.N. Ballantyne, R.I. Kemp, K.A. Martire, Corrigendum to “Communicating forensic science opinion: An examination of expert reporting practices” [*Science & Justice* 60 (3) (2020) 216–224], *Science & Justice* 61(4) (2021) 449–450. <https://doi.org/10.1016/j.scijus.2021.04.001>.
- [27] S. Garfinkel, P. Farrell, V. Roussev, G. Dinolt, Bringing science to digital forensics with standardized forensic corpora, DFRWS 2009, Montreal, Canada. <https://simso.n.net/clips/academic/2009.DFRWS.Corpora.pdf>, 2009 (accessed 4 July 2021).
- [28] G. Jackson, S. Jones, G. Booth, C. Champod, I.W. Evett, The nature of forensic science opinion - a possible framework to guide thinking and practice in investigations and in court proceedings, *Science & Justice* 46 (1) (2006) 33–44, [https://doi.org/10.1016/s1355-0306\(06\)71565-9](https://doi.org/10.1016/s1355-0306(06)71565-9).
- [29] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A hierarchy of propositions: deciding which level to address in casework, *Science & Justice* 4 (38) (1998) 231–239, [https://doi.org/10.1016/s1355-0306\(98\)72117-3](https://doi.org/10.1016/s1355-0306(98)72117-3).
- [30] W.C. Thompson, R.H. Grady, E. Lai, H.S. Stern, Perceived strength of forensic scientists' reporting statements about source conclusions, *Law, Probability & Risk* 17 (2) (2018) 133–155, <https://doi.org/10.1093/lpr/mgy012>.
- [31] W.C. Thompson, 2018. How should forensic scientists present source conclusions?, *Seton Hall Law Review* 48(3) Article 9. <https://scholarship.shu.edu/shlr/vol48/iss3/9>.
- [32] NAS (National Academy of Science), Strengthening forensic science in the United States: A path forward. <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf>, 2009 (accessed 4 July 2021).
- [33] PCAST (President's Council of Advisors on Science and Technology), Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods. Report to the President, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf, 2016 (accessed 4 July 2021).
- [34] AAAS (American Association of the Advancement of Science), Forensic science assessments: A quality and gap analysis – latent fingerprint examination (Report prepared by W. Thompson, J. Black, A. Jain, J. Kadane). https://www.aaas.org/sites/default/files/s3fs-public/reports/Latent%2520Fingerprint%2520Report%2520FINAL%25209_14.pdf, 2017 (accessed 4 July 2021).
- [35] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A model for case assessment and interpretation, *Science & Justice* 38 (3) (1998) 151–156, [https://doi.org/10.1016/s1355-0306\(98\)72099-4](https://doi.org/10.1016/s1355-0306(98)72099-4).
- [36] G. Jackson, *Understanding forensic science opinions*, in: J. Fraser, R. Williams (Eds.), *Handbook of forensic science*, Willian Publishing, Devon, 2009, pp. 419–445.
- [37] ENFSI, Best practice manual for the forensic examination of digital technology, ENFSI-BPM-FOT-01. Version 01 (November 2015). https://enfsi.eu/wp-content/uploads/2016/09/1_forensic_examination_of_digital_technology_0.pdf, 2015 (accessed 4 July 2021).
- [38] W.C. Thompson, E.L. Schumann. Interpretation of statistical evidence in criminal trials - the prosecutor's fallacy and the defence attorney's fallacy. *Law and Human Behavior*, 11(3) (1987) 167–187. <https://doi.org/10.1007/bf01044641>.
- [39] I.W. Evett, Avoiding the transposed conditional, *Science & Justice* 35 (2) (1995) 127–131, [https://doi.org/10.1016/s1355-0306\(95\)72645-4](https://doi.org/10.1016/s1355-0306(95)72645-4).
- [40] M. Lidén, Confirmation bias in criminal cases, Doctoral dissertation, Department of Law, Uppsala University, 2018. <http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Auu%3Adiva-351709> (accessed 4 July 2021).
- [41] D.M. Risinger, Reservations about likelihood ratios (and some other aspects of forensic 'Bayesianism'), *Law, Probability & Risk* 12 (1) (2013) 63–73, <https://doi.org/10.1093/lpr/mgs011>.
- [42] H. Henseler, C. de Poot, De betekenis van digitale sporen voor bewijs op activiteitsniveau, *Expertise en Recht*, (2) (2020) 50–59.
- [43] ENFSI, ENFSI guideline for evaluative reporting in forensic science. Strengthening the evaluation of forensic results across Europe (STEOFRAE). https://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf, 2015 (accessed 4 July 2021).
- [44] E. Casey, Error, uncertainty and loss in digital evidence, *International Journal of Digital Evidence*, 1(2) (2002).
- [45] E. Casey. *Digital evidence and computer crime: Forensic science, computers, and the internet*, Academic press, London, 2011.
- [46] G. Horsman, Digital evidence certainty descriptors (DECDS), *Forensic Science International: Digital Investigation* 32 (2020), 200896, <https://doi.org/10.1016/j.fsidi.2019.200896>.

Appendices:

1. Approval of application from the Norwegian Police Directorate
2. Approval of application from the Norwegian Attorney General
3. Norwegian centre for research data (NSD) approval
4. Information letter (Submission forms)
5. Information letter (DF experiment)
6. Part 1 survey
7. Information to participant on experiment day (control)
8. Information to participant on experiment day (weak guilt context)
9. Information to participant on experiment day (strong guilt context)
10. Information to participant on experiment day (innocence context)
11. Template – Log
12. Template – Analysis report
13. Part 2 survey
14. Co-author declaration – Article 1
15. Co-author declaration – Article 2
16. PW Allen Award – Article 5



Riksadvokaten
Postboks 2102 Vika
0125 Oslo

NATIONAL POLICE DIRECTORATE

Deres referanse:

Vår referanse:
 201803809-18 501

Sted, Dato
 Oslo, 08.08.2019

**VEDRØRENDE SØKNAD OM FRITAK FRA TAUSHETSPLIKTEN -
 FORSKNINGSPROSJEKT - A STUDY OF THE DIGITAL FORENSIC
 DETECTIVES ROLE IN THE CONSTRUCTION OF DIGITAL EVIDENCE IN
 CRIMINAL INVESTIGATION - NINA SUNDE**

Politidirektoratet viser til søknad av 28.08.2018 fra politioverbetjent Nina Sunde om fritak fra taushetsplikt for ansatte og særorganer og politidistrikter i forbindelse med doktorgradsprosjekt "A study of the digital forensic detectives' role in the construction of digital evidence in criminal investigation". Søknaden er oversendt både Politidirektoratet og Riksadvokaten.

Av søknaden fremkommer det at hovedfokus i doktorgradsprosjektet vil ligge på dataetterforskerens rolle i håndteringen av digitale bevis. Det opplyses at forskningsdata vil bli samlet inn gjennom intervju, dokumentanalyse, politirapporter og eksperiment. Søker ønsker å intervju ansatte på ulike nivåer i politiorganisasjonen som håndterer digitale bevis i sitt arbeid; herunder spesialister i Kripos, Økokrim, politidistriktene, samt ansatte i politiet uten spesialiststilling/kompetanse. I tilleggsinformasjon av 12.07.2019 går det frem at søker først vil gjennomgå straffesaksdokumentene og deretter forespørre 25-30 etterforskere/påtalejurister fra saker hun har fått tilgang til om å delta i intervju. Videre ønsker hun tillatelse til å gjennomføre et eksperiment med ca. 60 deltakere fra politiet. Deltakerne vil være dataetterforskere, både med sivil og politifaglig bakgrunn, og med dataetterforskning som sin hovedoppgave i politiet. Gjennom eksperimentet vil forskeren etter det opplyste få informasjon om hvordan politiet metodisk jobber i en analyse av et databaseslag og dokumentasjonen av dette arbeidet.

Politidirektoratets vurdering

Hva gjelder spørsmålet om det skal gis fritak for taushetsplikt for opplysninger til bruk for prosjektet som skisseres i søknaden, anser Politidirektoratet Riksadvokaten for å være rette vedkommende til å vurdere dette. Det vises til at opplysningene det ønskes tilgang til er innenfor påtalesporet og til at det bes om innsyn i straffesaksdokumenter i BL.

Politidirektoratet

Post: Postboks 2090 Vika, 0125 Oslo
 Besøk: Fridtjof Nansens vei 14/16

Tlf: 23 36 41 00
 Faks: 23 36 42 96
 E-post: politidirektoratet@politiet.no

Org. nr.: 982 531 950
 Giro: 7694.05.18020
 www.politi.no

Vi orienterer for øvrig om at søknaden er forelagt Kripos til vurdering. Slik Kripos leser søknaden er det ikke ønske om å få verken tilgang eller utlevert opplysninger fra noen av de sentrale registrene som Kripos er behandlingsansvarlig for. Søknaden er dermed ifølge Kripos ikke avhengig av godkjenning fra deres side.

Politidirektoratet anser for øvrig forskningen for å ha høy samfunnsmessig verdi og vurderer at prosjektet ikke reiser praktiske eller forskningsetiske problemstillinger av betydning. Politidirektoratet slutter seg imidlertid til Riksadvokatens foreløpige vurderinger av 11.07.2019 vedrørende søknadens omfang, hvor det uttales at den opprinnelige søknaden er for omfattende og for vanskelig å etterkomme, og at forsker må avgrense sin søknad. Riksadvokaten har bedt om at søknaden om tilgangen til straffesaker i BL må avgrenses både i tid og hva gjelder antall og type saker. Videre har Riksadvokaten bedt om opplysninger om hvordan utvelgelsen av intervjupersonene ønskes gjennomført.

Under forutsetning av at søknaden nedskaleres og avgrenses i henhold til Riksadvokatens vurderinger, godkjenner Politidirektoratet at det avsettes ressurser til gjennomføring av forskningen på de vilkår Riksadvokaten setter.

Politidirektoratet bemerker at forsker sendte endringssøknad i e-post av 12.07.2019.

Med hilsen

Kristine Langkaas
seksjonssjef

Amanda Baann Asdal
rådgiver

Dokumentet er elektronisk godkjent uten signatur.

Kopi til
Nina Sunde



RIKSADVOKATEN

Nina Sunde,
Politihøgskolen,
pb. 2109 Vika,
0125 Oslo

REF.:

VÅR REF.:

2018/01251-006 IW1001

DATO:

15.01.2020

SVAR PÅ SØKNAD OM INNSYN I STRAFFESAKER TIL FORSKNINGSFORMÅL

Det vises til søknad av 28.08 2018 om å benytte opplysninger fra straffesaker i Forskningsprosjektet "A study of the digital forensic detective's role in the construction of digital evidence in a criminal investigation", tilleggsskriv av 12.07. 2019 samt diverse samtaler om prosjektet.

Det fremgår av søknaden at prosjektet er et doktorgradsprosjekt ved Universitet i Oslo, Juridiske fakultet, Institutt for kriminologi og rettsosjologi under veiledning av professor Helene O. I. Gundhus (UIO), førsteamanuensis Johanne Yttri Dahl (PHS) og førsteamanuensis Fergas Thomas Tollen (PHS). Av søknaden fremkommer det at hovedfokuset i prosjektet vil ligge på dataetterforskernes rolle i håndteringen av digitale bevis i straffesaker. Det ønskes gjennom prosjektet å bidra til økt forskningsbasert kunnskap om politiets håndtering av digitale bevis samt kartlegge kilder til systematiske feil på individ og organisasjonsnivå. Forskningsdata ønskes innsamlet gjennom intervjuer, dokumentanalyser, politirapporter mv. Det ønskes videre å gjennomføre et eksperiment med deltakere fra politiet (jf. punkt 3 i søknaden) for å få informasjon om hvordan politiet metodisk jobber ved analyse av databaseslag og dokumentasjon av dette arbeidet.

Riksadvokaten anser at prosjektet faller inn under forskning og at det utvilsomt vil ha samfunnsnyttig verdi. Innsyn finnes rimelig og synes innen gitte rammer ikke å ville medføre uforholdsmessige ulempe for andre interesser. På denne bakgrunn beslutter riksadvokaten i medhold av politiregisterloven § 33 jf. 23 at søkeren gis tilgang til totalt 12 alvorlige straffesaker (10 års strafferamme) innen kategoriene vold, seksuallovbrudd, trusler og narkotika og 6 mindre alvorlige straffesaker innen kategoriene trusler, vold, vinning og vegtrafikk fra hvert politidistrikt. Videre gis søker innsyn i totalt 12 straffesaker fra Kripas innen datakriminalitet, menneskehandel og krigsforbrytelser og totalt 12 straffesaker fra Økokrim innen økonomisk kriminalitet og miljøkriminalitet, jf. søkers skriv av 12. 07. 2019. Det forutsettes at det i sakene ikke er benyttet skjulte

etterforskningsmetoder. Tilgangen omfatter både rettskraftig avgjorte saker og saker under arbeid i politiet. Sakene bør fortrinnsvis være fra de siste fem årene. Sakene velges ut av politidistriktene i samråd med søkeren og oversendes søkeren i PDF format enten på papir eller på passordbelagt/kryptert minnepinne. Der søker kun har behov for tilgang til enkelte dokumenter i sakene legges det til grunn at søkeren selv spesifiserer hvilke dokumenter dette er og at dette formidles til politidistriktet.

Som ledd i forskningsprosjektet har søker i brev 20.08. 2019 spesifikt bedt om innsyn i sak 109411744 (BL). Samtykke i innsyn omfatter også denne saken, som forutsettes gjort tilgjengelig for søkeren som øvrige saker nevnt foran.

Utvelgelsen av personer til intervju begrenses til 25-30 personer, jf. søkerens tilleggsskriv av 12.07.2019. Tjenestepersonene forutsettes å være blant personalet som har håndtert de straffesakene som søker er gitt tilgang til og at utvelgelsen skjer etter nærmere samråd med de enkelte tjenestepersonenes politidistrikt.

Det vises i søknaden til at det ønskes gjennomført et eksperiment (jf. punkt 3 i søknaden) Slik søknaden er utformet legges det til grunn at eksperimentet vil knyttes til en fiktiv sak og således ikke berøre taushetsbelagt informasjon.

Det forutsettes at all innsamling, oppbevaring og bruk av taushetsbelagt informasjon foregår på en faglig forsvarlig måte. Ved eventuell publikasjon av rapport eller annet må alle person identifiserte opplysninger anonymiseres og skjue innen rammene for forskeres taushetsplikt jf. politiregisterloven § 33 tredje ledd og henvisningen til forvaltningsloven § 13 e. Det forutsettes videre at alle person identifiserte opplysninger slettes fem år etter prosjektets slutt i 2024.

Tilgangen forutsetter at de i forskningsprosjektet som skal gis innsyn i opplysningene undertegner vedlagt skjema for erklæring om taushetsplikt og returnerer dette hit.

Vi beklager at det på grunn av et høyt arbeidspress ved embetet har tatt uforholdsmessig lang tid å behandle saken, og står til disposisjon i det videre arbeidet dersom vi kan bidra til fremdriften i prosjektet fremover.

[signature removed]

[signature removed]



RIKSADVOKATEN

ERKLÆRING OM TAUSHETSPLIKT

Jeg erkjenner å være gjort kjent med, mottatt kopi av og satt meg inn i taushetspliktsbestemmelsene i

- straffeprosessloven §§ 61a – 61c
- straffeprosessloven § 216i
- strafferegistreringsforskriftens § 20.

Jeg forplikter meg til å respektere ovennevnte bestemmelser, og således bevare taushet om og heller ikke på annen måte bidra til at uvedkommende får kjennskap til:

- personlige forhold om personer omhandlet i straffesaker
- opplysninger som bør holdes hemmelig av hensyn til etterforskningen av straffesaker
- enhver opplysning knyttet til saker om telefonkontroll
- enhver opplysning fra strafferegistrene.

Jeg er særlig gjort oppmerksom på at taushetsplikten gjelder etter at engasjementet ved Riksadvokatembetet er avsluttet, og at overtredelse av taushetspliktsbestemmelsene kan straffes etter straffeloven § 121 med bøter eller fengsel inntil 6 måneder.

Oslo, 22.1.20

[signature removed]

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD sin vurdering

Prosjekttittel

A study of the digital forensic detectives' role in the construction of digital evidence in criminal investigation

Referansenummer

458568

Registrert

14.08.2019 av Nina Sunde - Nina.Sunde@phs.no

Behandlingsansvarlig institusjon

Universitetet i Oslo / Det juridiske fakultet / Institutt for kriminologi og rettssosiologi

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Helene O. I. Gundhus, h.o.i.gundhus@jus.uio.no, tlf: 41523351

Felles behandlingsansvarlige institusjoner

Politihøgskolen

Type prosjekt

Forskerprosjekt

Prosjektperiode

14.08.2018 - 14.08.2024

Status

12.09.2019 - Vurdert

Vurdering (1)

12.09.2019 - Vurdert

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 12.09.2019, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

MELD VESENTLIGE ENDRINGER

Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde:

https://nsd.no/personvernombud/meld_prosjekt/meld_endringer.html

Du må vente på svar fra NSD før endringen gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET

Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 14.08.2024.

LOVLIG GRUNNLAG

Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake. Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

For utvalg 2 skal det først innhentes politirapporter (med unntak av navnet til rapportskriver inneholder disse ingen personopplysninger ved utlevering) og rekrutteringen gjøres med bakgrunn i disse rapportene og de utvalgte blir spurt om de ønsker å delta i intervju hvortil det innhentes samtykke. For denne datakilden forutsetter vi at den innestående søknaden for dispensasjon av taushetsplikten for utlevering av disse godkjennes. Hvis ikke må denne datakilden utelukkes.

PERSONVERNPRINSIPPER

NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER

Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

NSD vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER

NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

Politihøgskolen er felles behandlingsansvarlig institusjon. NSD legger til grunn at behandlingen oppfyller kravene til felles behandlingsansvar, jf. personvernforordningen art. 26.

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET

NSD vil følge opp underveis (hvert annet år) og ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet/pågår i tråd med den behandlingen som er dokumentert.

Lykke til med prosjektet!

Kontaktperson hos NSD: Karin Lillevold
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

Dear XX

I am contacting you because, in relation to my PhD project, we are doing research to better understand and improve digital forensic work. As a first step, we are interested in how work is commissioned by police or other entities. To this end, we are collecting submission forms, to explore how assignments from digital forensic examiners are requested. This is to develop a better knowledge foundation about how mandates or assignments for digital investigation/digital forensics are conveyed.

In this regard, we are asking for your help in gathering forms from different law enforcement units or private laboratories/units throughout the world assisting in digital forensics/digital investigation. We therefore would appreciate if you would confidentially and anonymously share the form used by our unit.

If you are not using a standard form, it would be of interest to us to know how digital forensic work in your unit is commissioned. If you do not use a form, then we would therefore appreciate if you could give a brief explanation on how the tasks/assignments for the digital forensic examiner are requested and agreed upon.

In the dissemination of our research (research articles, presentations) we will never disclose or share your form, nor identifiable information such as your name or organisation. So your participation and the form is totally confidential and anonymous. We will aggregate forms across agencies and only report statistical information. An example of what we might convey is e.g., "X% of the forms provided a 'priority' designation, while the rest did not specify the priority of the work. From those forms that specified a priority level, Y% designated the priority as 'low', 'medium' and 'high', and Z% used a scale of 1-10". Personal data, such as your name and email address are only used for correspondence and collection of forms, and will not be stored in the project.

We very much appreciate your help and cooperation, and looking forward to receiving a copy of the submission form used in your lab, or, information that no form is used.

If you have any questions, please do not hesitate to ask.

Best regards

Nina Sunde

Police Superintendent

Department for post graduate education

The Norwegian Police University College

Phone: +47 91660069 / +47 23199583

Would you like to participate in the research project "The digital forensic detective's role in the construction of digital evidence in a criminal investigation" ?

We would like to invite you to participate in a research project where the purpose is to explore the digital forensic examiners role in the construction of digital evidence. In this information letter, we will provide information about the purpose for the project, and what participation entails for you.

Purpose

The project is part of a doctoral study, with the purpose of developing more knowledge about how digital evidence is handled by law enforcement, and what role the different professional actors play during a criminal investigation. The main focus is the digital forensic examiner.

Who is responsible for the research project?

The responsible institutions for the project are The Norwegian Police University College (NPUC) and the University of Oslo (UiO).

The project is conducted by:

Nina Sunde,
Detective Superintendent and PhD student, NPUC / UiO
Politihøgskolen (NPUC), PB. 5027 Majorstuen, 0301 Oslo.
E-mail: nina.sunde@phs.no, phone: + 47 91 66 00 69

Main supervisor:

Helene O.I. Gundhus (UiO) – *Responsible for the project*

About the participation

The project is mainly aimed at digital forensic examiners who have digital forensic work as their primary task. I would like to invite you to participate in an experiment related to this project. The experiment takes 4-5 hours to complete. Participation will involve analysing a fictitious evidence file, and writing a report about the analysis. You will also be asked to fill out a short questionnaire prior to and after the analysis. The first questionnaire is concerned about background information (age, education, experience etc.), and the last is about your judgements and decisions during the analysis of the evidence file. The report you write, as well as the answers in the questionnaire, will be anonymised, and saved in the research project. You will not be asked for information for which you have a duty of confidentiality, such as information on real criminal cases.

Participation in the project is voluntary. If you choose to participate, you may withdraw your consent at any time without giving any reason. All information about you will then be anonymised. It will not have any negative consequences for you if you do not want to participate or later choose to withdraw.

Personal data

Your personal information will only be used for the purpose stated in this letter. All information will be handled with confidentiality, and in compliance with privacy regulations. Reports and questionnaires will be stored on an encrypted and password protected hard drive, and will only be accessible to the persons responsible for the project. Your name and contact information will be replaced with a code that is stored on a list which is stored separately, and away from the other data. This list will only be available to Sunde and Gundhus. In dissemination of the research, we will not disclose any information that may be traced back to individual participants.

What happens to your personal information when the research project ends?

The PhD project is planned to be due 14.08.2024. When the research project ends, your personal information, the reports and answers in the questionnaires will be deleted.

Your rights (obligatory information for you, due to the Personal Data Act)

While you may be identified within the data material, you have the right to:

- insight into what personal data is registered about you,
- to have your personal information corrected,
- get personal information about you deleted,
- get a copy of your personal data (data portability), and
- to submit a complaint to the Privacy Ombudsman or the Norwegian Data Protection Authority regarding the processing of your personal data.

Why may we handle personal data about you?

We may handle personal data about you based on your consent. The Norwegian Centre for Research Data have evaluated the handling of personal data in this project, and concluded that they are handled in compliance with privacy regulations.

If you have any questions about this research, or wish to exercise your rights, please contact:

- Nina Sunde (NPUC) or Helene Gundhus (UiO).
- Our Privacy Ombudsman: Knut Erik Hauslo, (NPUC) – phone: +47 23 19 99 00
- The Norwegian Centre for Research Data AS, e-mail: (personverntjenester@nsd.no) or phone: + 47 55 58 21 17.

Confidentiality

Since all participants will not complete the experiment during the same day, the project will be run over a period of time. It is therefore of great importance that you do not share information about the experiment to others until the results are published, such as the scenario from which the evidence file is related, your findings, your opinions about the findings, your answers to the questionnaire or your analysis report. Such information may affect how new participants conduct the experiment and will be very unfortunate for the result.

Kind regards

Helene Gundhus
Project responsible and supervisor

Nina Sunde
PhD student

Declaration of consent

I have received and understood information about *the Digital Forensic Detective's role in the construction of digital evidence in a criminal investigation*, and have had the opportunity to ask questions. I agree to:

- Participate in the experiment
- Not to share information to others about the details of the experiment

I agree that my information will be processed until the project is completed, approximately 14th of August, 2024.

(Signed by participant, date)

Part 1 Background information

Please note: 2 pages

0. Gender

- Male ()
- Female ()

1. Education (tick off one or both options):

- Civil ()
- Police ()

2. Level of education (tick off only one option):

- Bachelor ()
- Master ()
- Ph.D. ()

3. Highest level of post graduate education in the Nordic NCFI (Nordic Computer Forensic Investigator) programme (tick off only one option):

- NCFI Introduction, 5 ECTS ()
- NCFI Core Concepts 15 ECTS ()
- NCFI module 2 (Advanced computer forensics/Online investigation/Network forensics and cybercrime), 15 ECTS ()
- NCFI module 3 (Forensic tool development/Linux artefacts/Linux as an investigative platform/Macintosh computer forensics/Windows forensics), 7,5 ECTS ()
- None of these ()

4. Other post graduate education in criminal investigation (you may tick off more options):

- Investigation methodology (general) ()
- Investigation of sexual crimes ()
- Investigation of violent crimes ()
- Investigation of organised crime ()
- Other post graduate educations within criminal investigation: (you may specify below)

5. In which level is your position (tick off one option):

- Federal level / National level ()
- State/territory level (within a state/territory centralised unit) ()
- Local level (within a police district unit) ()

(B) Part 1 Background information

6. Experience (years) :

- Years of experience within law enforcement: (___)
- Years of experience with criminal investigations: (___)
- Years of experience as digital forensic examiner within law enforcement: (___)
- Years of experience as digital forensic examiner outside law enforcement: (___)

7. Analysis software:

**Which analysis software do you use the most of the following
(rank from most used: 1, to least used: 4. If you are not using the software, write X)**

- EnCase (Guidance Software)_____
- X-Ways (X-Ways Forensics)_____
- Axiom (Magnet Forensics)_____
- Forensic Toolkit FTK (AccessData)_____
- Other (please specify):_____

Description of the experiment

Thank you for participating.

By now, you have completed and returned:

- (A) Information letter and consent form
- (B) Part 1 – Background information

It is now time to analyse an evidence file from the following case:

Confidential information leakage

M57.biz is a small US based company, with office in your country. The company, which develops and sells body art equipment (tattoo, piercing etc.), is in the start-up phase. The manager for the M57.biz office in your country is Alison Smith, and the CFO is Jean Jones. The company has 4 programmers, 2 in marketing, and 1 in business development. Only Alison and Jean have a permanent office space, while the other employees work from home office. All employees participate in a daily online meeting. There are in-person meetings for all employees in the M57.biz office once every two weeks. Most documents are exchanged by email.

A spreadsheet (m57plan.xls) containing confidential information was recently posted as an attachment in a forum of a competitor's website. When this was discovered, Alison reported the incident to the police as information theft. Alison told police that Jean, the CFO, was responsible for updating the spreadsheet, and that it was probably sent from Jean's computer.

The attachment posted on the competitor's website looked like this:

M57.biz company				
Name		Position	Salary	SSN (for background check)
Allison	Smith	President	\$140,000	103-44-3134
Jean	Jones	CFO	\$120,000	432-34-6432
Programmers:				
Bob	Blackman	Apps 1	90,000	493-46-3329
Carol	Canfred	Apps 2	110,000	894-33-4560
Dave	Daubert	Q&A	67,000	331-95-1020
Emmy	Arlington	Entry Level	57,000	404-98-4079
Marketing				
Gina	Tangers	Creative 1	80,000	980-97-3311
Harris	Jenkins	G & C	105,000	887-33-5532
BizDev				
Indy	Counterching	Outreach	240,000	123-45-6789
Annual Salaries			\$1,009,000	
Benefits			30%	\$302,700

You are tasked with analysing a copy of the hard drive from Jeans computer, and find out:

What has happened, and what was Jeans involvement in the reported incident?

NB: This is an evidence file with timestamps from 2008. However, you should imagine that you are investigating the case immediately after it happened, i.e. in 2008 and not today (2020).

How to conduct the assignment:

You are tasked with analysing the evidence file. Please use Axiom (Magnet Forensics) and / or X-Ways (X-Ways Forensics) if they are familiar and available to you. If not, you can use your preferred analysis software. While analysing, keep a log using the attached template (C - Log). Bookmark information of interest during the analysis. When the analysis is over, export the bookmarks to a PDF file. Then you write analysis report where the results are documented in the way you would normally do it, and use the attached template (E - Analysis report). The whole assignment (both the analysis and the report writing) must be done individually, and it is very important that you do not consult or confer with others during the experiment.

The output of the experiment that should be handed in is:

- (C) Log (please use the received template)
- (D) PDF with exported bookmarks
- (E) Report from analysis (please use the received template)

When you have handed in C, D and E, you will receive the final part on e-mail:

(F) Part 2 – Final questions

When part F is completed, you hand it in by e-mail, and the experiment is completed.

Remember: It is important not to share information with others about your experiment, such as what you discovered, what you noted in the log, or what you wrote in your report. This can affect how new participants conduct the experiment, and will be very unfortunate for the result.

Description of the experiment

Thank you for participating.

By now, you have completed and returned:

- (A) Information letter and consent form
- (B) Part 1 – Background information

It is now time to analyse an evidence file from the following case:

Confidential information leakage

M57.biz is a small US based company, with office in your country. The company, which develops and sells body art equipment (tattoo, piercing etc.), is in the start-up phase. The manager for the M57.biz office in your country is Alison Smith, and the CFO is Jean Jones. The company has 4 programmers, 2 in marketing, and 1 in business development. Only Alison and Jean have a permanent office space, while the other employees work from home office. All employees participate in a daily online meeting. There are in-person meetings for all employees in the M57.biz office once every two weeks. Most documents are exchanged by email.

A spreadsheet (m57plan.xls) containing confidential information was recently posted as an attachment in a forum of a competitor's website. When it was discovered, Alison reported the incident to the police as information theft. Alison told police that Jean, the CFO, was responsible for updating the spreadsheet, and that it was probably sent from Jean's computer. Alison told police that there has been a long-lasting wage dispute with the programmers in the firm, who claim to be underpaid. One of the programmers - Emmy Tuckford Arlington – has fronted the dispute on behalf of the programmers in M57.biz. Jean has supported the programmers in this conflict, and has told Alison that the company can afford to pay them better salaries.

Jean is about to be interviewed by the police about the reported incident. However, the chief investigating officer wants an analysis of Jean's computer before the police interview, to look for traces indicating that she was involved in the reported incident.

The attachment posted on the competitor's website looked like this:

M57.biz company				
Name		Position	Salary	SSN (for background check)
Allison	Smith	President	\$140,000	103-44-3134
Jean	Jones	CFO	\$120,000	432-34-6432
Programmers:				
Bob	Blackman	Apps 1	90,000	493-46-3329
Carol	Canfred	Apps 2	110,000	894-33-4560
Dave	Daubert	Q&A	67,000	331-95-1020
Emmy	Arlington	Entry Level	57,000	404-98-4079
Marketing				
Gina	Tangers	Creative 1	80,000	980-97-3311
Harris	Jenkins	G & C	105,000	887-33-5532
BizDev				
Indy	Counterchng	Outreach	240,000	123-45-6789
Annual Salaries			\$1,009,000	
Benefits			30%	\$302,700

You are tasked with analysing a copy of the hard drive from Jeans computer, and find out:

What has happened, and what was Jeans involvement in the reported incident?

NB: This is an evidence file with timestamps from 2008. However, you should imagine that you are investigating the case immediately after it happened, i.e. in 2008 and not today (2020).

How to conduct the assignment:

You are tasked with analysing the evidence file. Please use Axiom (Magnet Forensics) and / or X-Ways (X-Ways Forensics) if they are familiar and available to you. If not, you can use your preferred analysis software. While analysing, keep a log using the attached template (C - Log). Bookmark information of interest during the analysis. When the analysis is over, export the bookmarks to a PDF file. Then you write analysis report where the results are documented in the way you would normally do it, and use the attached template (E - Analysis report). The whole assignment (both the analysis and the report writing) must be done individually, and it is very important that you do not consult or confer with others during the experiment.

The output of the experiment that should be handed in is:

- (C) Log (please use the received template)
- (D) PDF with exported bookmarks
- (E) Report from analysis (please use the received template)

When you have handed in C, D and E, you will receive the final part on e-mail:

- (F) Part 2 – Final questions

When part F is completed, you hand it in by e-mail, and the experiment is completed.

Remember: It is important not to share information with others about your experiment, such as what you discovered, what you noted in the log, or what you wrote in your report. This can affect how new participants conduct the experiment, and will be very unfortunate for the result.

Description of the experiment

Thank you for participating.

By now, you have completed and returned:

- (A) Information letter and consent form
- (B) Part 1 – Background information

It is now time to analyse an evidence file from the following case:

Confidential information leakage

M57.biz is a small US based company, with office in your country. The company, which develops and sells body art equipment (tattoo, piercing etc.), is in the start-up phase. The manager for the M57.biz office in your country is Alison Smith, and the CFO is Jean Jones. The company has 4 programmers, 2 in marketing, and 1 in business development. Only Alison and Jean have a permanent office space, while the other employees work from home office. All employees participate in a daily online meeting. There are in-person meetings for all employees in the M57.biz office once every two weeks. Most documents are exchanged by email.

A spreadsheet (m57plan.xls) containing confidential information was recently posted as an attachment in a forum of a competitor's website. When this was discovered, Alison reported the incident to the police as information theft. Alison told police that Jean, the CFO, was responsible for updating the spreadsheet, and that it was probably sent from Jean's computer. Jean was arrested for information theft, and in a police interview earlier today, she confessed that she had committed the criminal offense.

The attachment posted on the competitor's website looked like this:

M57.biz company				
Name		Position	Salary	SSN (for background check)
Allison	Smith	President	\$140,000	103-44-3134
Jean	Jones	CFO	\$120,000	432-34-6432
Programmers:				
Bob	Blackman	Apps 1	90,000	493-46-3329
Carol	Canfred	Apps 2	110,000	894-33-4560
Dave	Daubert	Q&A	67,000	331-95-1020
Emmy	Arlington	Entry Level	57,000	404-98-4079
Marketing				
Gina	Tangers	Creative 1	80,000	980-97-3311
Harris	Jenkins	G & C	105,000	887-33-5532
BizDev				
Indy	Counterchng	Outreach	240,000	123-45-6789
Annual Salaries			\$1,009,000	
Benefits			30%	\$302,700

You are tasked with analysing a copy of the hard drive from Jeans computer, and find out:

What has happened, and what was Jeans involvement in the reported incident?

NB: This is an evidence file with timestamps from 2008. However, you should imagine that you are investigating the case immediately after it happened, i.e. in 2008 and not today (2020).

How to conduct the assignment:

You are tasked with analysing the evidence file. Please use Axiom (Magnet Forensics) and / or X-Ways (X-Ways Forensics) if they are familiar and available to you. If not, you can use your preferred analysis software. While analysing, keep a log using the attached template (C - Log). Bookmark information of interest during the analysis. When the analysis is over, export the bookmarks to a PDF file. Then you write analysis report where the results are documented in the way you would normally do it, and use the attached template (E - Analysis report). The whole assignment (both the analysis and the report writing) must be done individually, and it is very important that you do not consult or confer with others during the experiment.

The output of the experiment that should be handed in is:

- (C) Log (please use the received template)
- (D) PDF with exported bookmarks
- (E) Report from analysis (please use the received template)

When you have handed in C, D and E, you will receive the final part on e-mail:

- (F) Part 2 – Final questions

When part F is completed, you hand it in by e-mail, and the experiment is completed.

Remember: It is important not to share information with others about your experiment, such as what you discovered, what you noted in the log, or what you wrote in your report. This can affect how new participants conduct the experiment, and will be very unfortunate for the result.

Description of the experiment

Thank you for participating.

By now, you have completed and returned:

- (A) Information letter and consent form
- (B) Part 1 – Background information

It is now time to analyse an evidence file from the following case:

Confidential information leakage

M57.biz is a small US based company, with office in your country. The company, which develops and sells body art equipment (tattoo, piercing etc.), is in the start-up phase. The manager for the M57.biz office in your country is Alison Smith, and the CFO is Jean Jones. The company has 4 programmers, 2 in marketing, and 1 in business development. Only Alison and Jean have a permanent office space, while the other employees work from home office. All employees participate in a daily online meeting. There are in-person meetings for all employees in the M57.biz office once every two weeks. Most documents are exchanged by email.

A spreadsheet (m57plan.xls) containing confidential information was recently posted as an attachment in a forum of a competitor's website. When this was discovered, Alison reported the incident to the police as information theft. Alison told police that Jean, the CFO, was responsible for updating the spreadsheet, and that it was probably sent from Jean's computer. As a result, Jean was arrested for information theft and questioned about the incident in a police interview. However, after the police interview, the police believe she is innocent, and that she was framed during a phishing attack.

The attachment posted on the competitor's website looked like this:

M57.biz company				
Name		Position	Salary	SSN (for background check)
Allison	Smith	President	\$140,000	103-44-3134
Jean	Jones	CFO	\$120,000	432-34-6432
Programmers:				
Bob	Blackman	Apps 1	90,000	493-46-3329
Carol	Canfred	Apps 2	110,000	894-33-4560
Dave	Daubert	Q&A	67,000	331-95-1020
Emmy	Arlington	Entry Level	57,000	404-98-4079
Marketing				
Gina	Tangers	Creative 1	80,000	980-97-3311
Harris	Jenkins	G & C	105,000	887-33-5532
BizDev				
Indy	Counterchng	Outreach	240,000	123-45-6789
Annual Salaries			\$1,009,000	
Benefits			30%	\$302,700

You are tasked with analysing a copy of the hard drive from Jeans computer, and find out:

What has happened, and what was Jeans involvement in the reported incident?

NB: This is an evidence file with timestamps from 2008. However, you should imagine that you are investigating the case immediately after it happened, i.e. in 2008 and not today (2020).

How to conduct the assignment:

You are tasked with analysing the evidence file. Please use Axiom (Magnet Forensics) and / or X-Ways (X-Ways Forensics) if they are familiar and available to you. If not, you can use your preferred analysis software. While analysing, keep a log using the attached template (C - Log). Bookmark information of interest during the analysis. When the analysis is over, export the bookmarks to a PDF file. Then you write analysis report where the results are documented in the way you would normally do it, and use the attached template (E - Analysis report). The whole assignment (both the analysis and the report writing) must be done individually, and it is very important that you do not consult or confer with others during the experiment.

The output of the experiment that should be handed in is:

- (C) Log (please use the received template)
- (D) PDF with exported bookmarks
- (E) Report from analysis (please use the received template)

When you have handed in C, D and E, you will receive the final part on e-mail:

- (F) Part 2 – Final questions

When part F is completed, you hand it in by e-mail, and the experiment is completed.

Remember: It is important not to share information with others about your experiment, such as what you discovered, what you noted in the log, or what you wrote in your report. This can affect how new participants conduct the experiment, and will be very unfortunate for the result.

(C) Log

Log

This document should be used during your analysis of the evidence file. Please make notes about what you are examining (e.g., review of the email, keyword search, etc.), why, what was the result, whether you bookmarked the result, and a brief assessment of the evidential value of the result in relation to the case under investigation. In the first line, enter the time when you started the analysis, and in the last line, when you finished.

Actions/what was examined	Justification/purpose	Result	Bookmarked Yes/no	Assessment of result
Start, date and time:				
Finished, date and time:				

(To get more lines in the table, place the cursor in the last cell, and press tab)

Report from analysis of evidence file

M57.biz

Author:

Date:

Part 2- Final questions

Part 2 contains 9 questions. Questions 1-7 is obligatory. Question 8 should not be answered if you have not made any findings beyond what is stated in question 7 . Question 9 is open ended, and may be answered if you wish to comment on any of your answers from questions 7 or 8.

1. Was the case M57.biz known to you from before?

No ()

Yes ()

If yes, please specify:

2. Which analysis software did you use? (you may tick of several options)

Axiom () Version:

X-Ways () Version:

EnCase () Version:

Forensic Tool Kit () Version:

Others? (if yes, please specify - name and version)

Answer:

3. After reading the introduction about the case, and prior to the analysis – what did you think had happened in relation to the reported incident?

Answer:

4. After the analysis, and documentation of the findings in the analysys report, what did you think had happened in relation to the reported incident?

Answer:

5. Did you use any techniques to safeguard your objectivity during the analysis? If yes, please specify.

(F) Part 2 Final questions

Answer:

6. Did you use any techniques to examine or control the reliability of the evidence during the analysis? If yes, please specify.

Answer:

7. Here you should state what information you found, and how you evaluate it in relation to Jean's guilt/innocence. You should place only one X per row. You must indicate whether the information indicates / substantiates that Jean is guilty, or whether it indicates / substantiates that Jean is innocent, or whether it is ambiguous (i.e. neither), or whether you did not find the relevant information during your analysis.

Time stamps are in UTC + 0:00

	Information	The information indicates /substantiates that Jean is guilty of a crime	The information indicates /substantiates that Jean is innocent	The information is ambiguous	Did not find this piece of information
1	The spreadsheet m57biz.xls (In the Desktop folder)				
2	AIM chat log between alisonm57 og Jean 18th Juli				
3	Email exchange between Jean and alison@m57.biz 19th and 20th July				
4	Email exchange between Jean and alex@m57.biz 19th and 20th July				
5	Email exchange between Jean and tuckgorge@gmail.com 20th of July				
6	Email exchange between Jean and bob@m57.biz 20th and 21st July				
7	Email exchange between Jean and carol@m57.biz 20th and 21st July				
8	Installation of QQGames Bubble Arena				

(F) Part 2 Final questions

	18th July				
9	Mounting of USB flash 18th July				
10	Creation of user profile Devon				
11	Creation of user profile Administrator				
12	Running of the program outlook.exe by Administrator				
13	Running of the program firefox.exe by Administrator				
14	Running of the program aim6.exe by Administrator				
15	Running of the program winword.exe by Administrator				
16	Running of the program notepad.exe by Administrator				
17	Running of the program cmd.exe by Administrator				

8. If you found any other important information that you think is relevant to the issue of Jean's guilt / innocence, you can fill it in the table below: (press tab in the last cell if you need more lines)

	Information	The information indicates /substantiates that Jean is guilty of a crime	The information indicates /substantiates that Jean is innocent	The information is ambiguous	
18					
19					
Etc.					

9. Other comments or remarks?

Answer:

Co-author declaration

Describing the independent research contribution of the candidate and each co-author

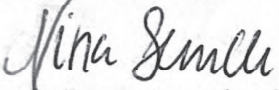
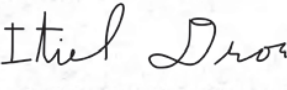
With reference to the Regulations for the degree of Philosophiae Doctor (PhD) at the University of Oslo, cf. Section 10.1: "Theses that include works by several authors shall be accompanied by a signed declaration describing the contributions made to each work by the candidate and each individual co-author."

The co-author declaration must be filled in electronically and signed by the candidate and co-author. Only the five most important co-authors of an article have to sign the declaration. Each co-author must complete one co-author declaration. The candidate must sign each co-author declaration and must make sure that the declaration and signatures are on the same page.

NB! The candidate must enclose the co-author declaration(s) with his/her application for thesis evaluation.

<p>Article no. : <u>1</u></p> <p>Title of article: <u>Cognitive and Human Factors in Digital Forensics: Problems, Challenges, and the Way Forward</u></p> <p>Name of candidate: <u>Nina Sunde</u></p> <p>First author: <input checked="" type="checkbox"/> Shared first authorship: <input type="checkbox"/> Second author: <input type="checkbox"/> Senior author: <input type="checkbox"/> Other: <input type="checkbox"/></p> <p>The independent contribution of the candidate: <u>Provided version 1, 3, 5 and 7 of the article</u></p> <hr/> <p>To the best of your knowledge, has this article been part of a previously evaluated doctoral thesis? Yes: <input type="checkbox"/> / No: <input checked="" type="checkbox"/></p> <p>If yes, please elaborate: _____</p> <hr/> <p>Do you know if one of your co-authors is going to use this article in his/her doctoral thesis? Yes: <input type="checkbox"/> / No: <input checked="" type="checkbox"/></p> <p>If yes, please name the co-author: _____</p>
--

<p>Co-author: <u>Itiel E. Dror</u></p> <p>First author: <input type="checkbox"/> Shared first authorship: <input type="checkbox"/> Second author: <input checked="" type="checkbox"/> Senior author: <input type="checkbox"/> Other: <input type="checkbox"/></p> <p>The independent contribution of the co-author: <u>Provided version 2, 4 and 6 of the article</u></p> <hr/>

<p>Must be signed by the candidate and co-author</p>	
<p></p> <p>Handwritten signature of candidate</p>	<p></p> <p>Handwritten signature of co-author</p>

Co-author declaration

Describing the independent research contribution of the candidate and each co-author

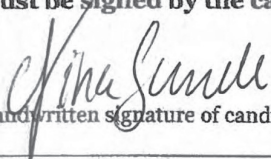
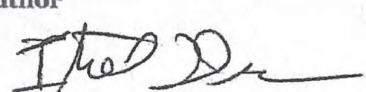
With reference to the Regulations for the degree of Philosophiae Doctor (PhD) at the University of Oslo, cf. Section 10.1: "Theses that include works by several authors shall be accompanied by a signed declaration describing the contributions made to each work by the candidate and each individual co-author."

The co-author declaration must be filled in electronically and signed by the candidate and co-author. Only the five most important co-authors of an article have to sign the declaration. Each co-author must complete one co-author declaration. The candidate must sign each co-author declaration and must make sure that the declaration and signatures are on the same page.

NB! The candidate must enclose the co-author declaration(s) with his/her application for thesis evaluation.

<p>Article no. : <u>301175</u></p> <p>Title of article: <small>A hierarchy of expert performance (HEP) applied to digital forensics. Reliability and biasability in digital forensics decision making</small> <u></u></p> <p>Name of candidate: <u>Nina Sunde</u></p> <p>First author: <input checked="" type="checkbox"/> Shared first authorship: <input type="checkbox"/> Second author: <input type="checkbox"/> Senior author: <input type="checkbox"/> Other: <input type="checkbox"/></p> <p>The independent contribution of the candidate: <u>Idea, project design, planning, recruiting participants, conducting pre-study (submission forms) and quasi-experiment, analysis, writing the research paper</u></p> <p>To the best of your knowledge, has this article been part of a previously evaluated doctoral thesis? Yes: <input type="checkbox"/> No: <input checked="" type="checkbox"/></p> <p>If yes, please elaborate: <u></u></p> <p>Do you know if one of your co-authors is going to use this article in his/her doctoral thesis? Yes: <input type="checkbox"/> No: <input checked="" type="checkbox"/></p> <p>If yes, please name the co-author: <u></u></p>
--

<p>Co-author: <u>Itiel E. Dror</u></p> <p>First author: <input type="checkbox"/> Shared first authorship: <input type="checkbox"/> Second author: <input checked="" type="checkbox"/> Senior author: <input type="checkbox"/> Other: <input type="checkbox"/></p> <p>The independent contribution of the co-author: <u>input on project design, collection for pre-study supervision on analysis/presentation of results, collaboration on writing/revision of the research paper.</u></p>
--

<p>Must be signed by the candidate and co-author</p>	
<p></p> <p>Handwritten signature of candidate</p>	<p></p> <p>Handwritten signature of co-author</p>



The
Chartered
Society of
Forensic
Sciences

PW Allen Award for the Most Meritorious Research Paper published in Science & Justice

“What does a digital forensics opinion look like? A comparative study of digital forensics and forensic science reporting practices”

Nina Sunde

for the year

2021

President

Honorary Secretary