Maastricht University

Facilitating falsification in legal decision making

Citation for published version (APA):

Maegherman, E. F. L. (2021). *Facilitating falsification in legal decision making: problems in practice and potential solutions*. [Doctoral Thesis, Maastricht University, University of Gothenburg]. The House of Legal Psychology. https://doi.org/10.26481/dis.20210114em

Document status and date:
Published: 01/01/2021

DOI:
10.26481/dis.20210114em

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

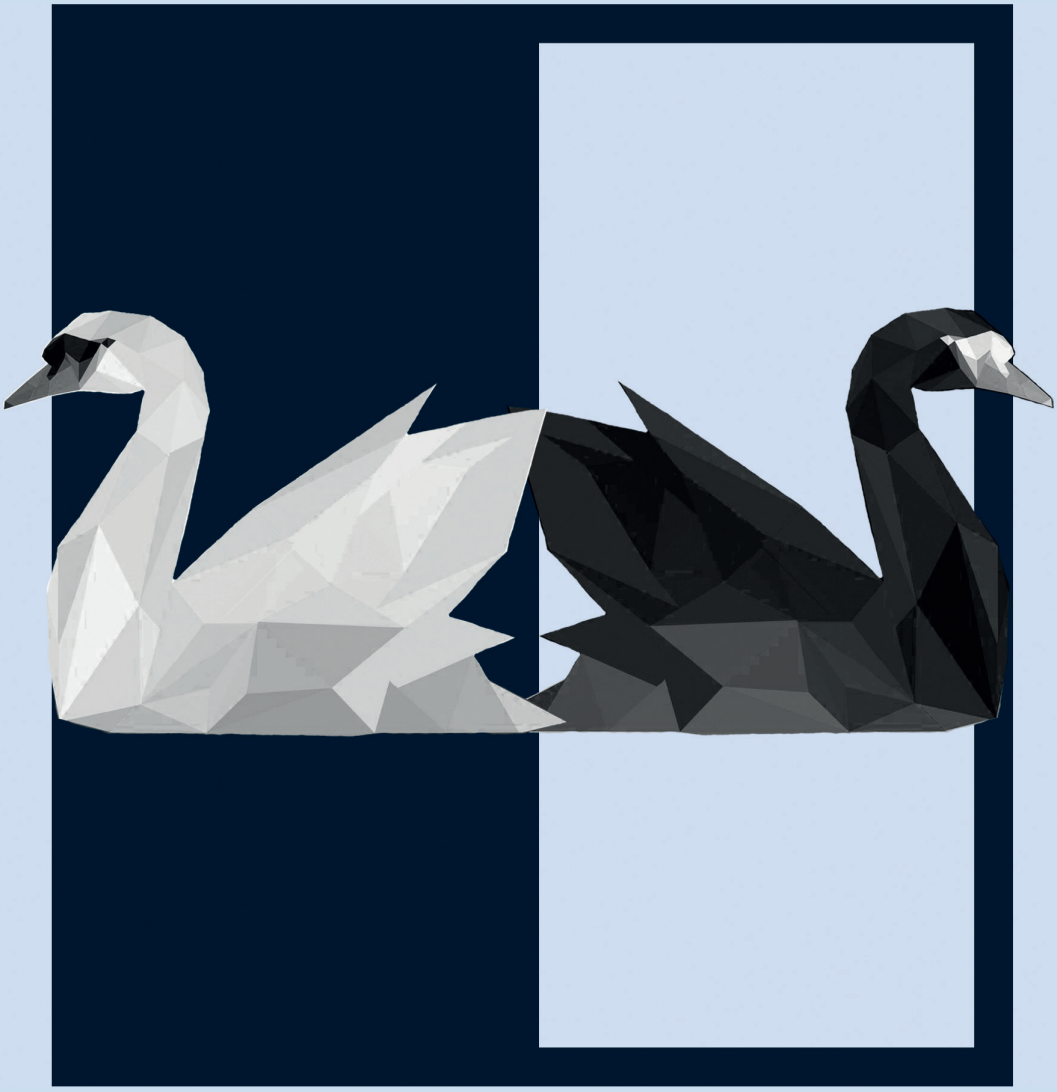
Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 19 Apr. 2024

FACILITATING FALSIFICATION IN LEGAL DECISION-MAKING: PROBLEMS IN PRACTICE AND POTENTIAL SOLUTIONS



ENIDE F. L. MAEGHERMAN

FACILITATING FALSIFICATION IN LEGAL DECISION-MAKING: PROBLEMS IN PRACTICE AND POTENTIAL SOLUTIONS

ENIDE F. L. MAEGHERMAN

© Enide Maegherman, 2020

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means without prior permission in writing by the author.

Cover Design: Laura L. E. Baker

Layout and design: Sanne Kassenberg, persoonlijkproefschrift.nl

Printing: Ridderprint | www.ridderprint.nl

ISBN: 978-94-6416-227-1

Maastricht University

UNIVERSITY OF GOTHENBURG

Facilitating Falsification in Legal Decision-Making: Problems in Practice and Potential Solutions

DISSERTATION

To obtain the degree of Doctor of Philosophy by the University of Gothenburg, on the authority of the Deputy Vice-Chancellor dr. Mete Sandoff, in accordance with the decision of the Faculty Board of Social Sciences and the degree of Doctor at the Maastricht University, on the authority of the Rector Magnifcus, Prof. dr. Rianne M. Letschert, in accordance with the decision of the Board of Deans, to be defended in public on
Thursday 14th January 2021 at 14.00

by

Enide Felten Lidewij Maegherman

The House of Legal Psychology

Supervisors:

Prof. dr. P.J. van Koppen	Maastricht University and VU University Amsterdam
Prof. dr. K. Ask	University of Gothenburg

Co-supervisor:

Dr. R. Horselenberg	Maastricht University

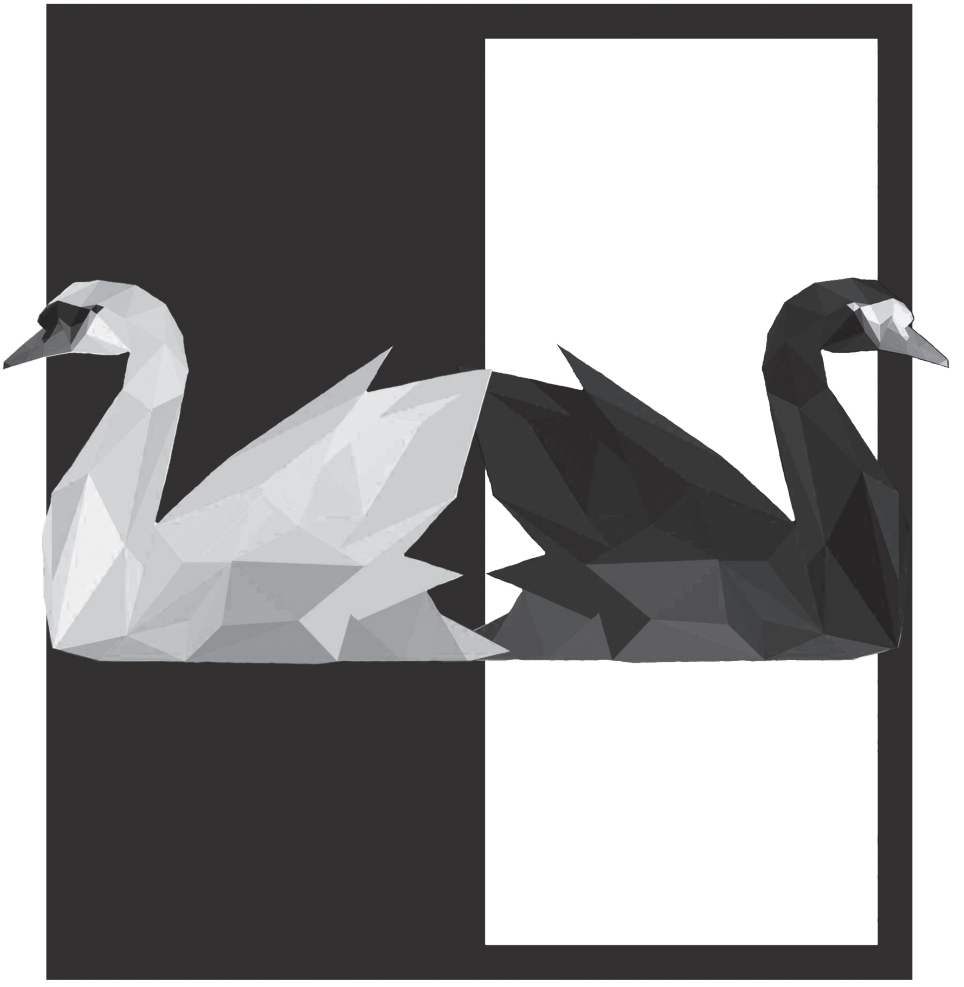
Assessment Committee

Prof. dr. D.V.A. Brouwer (chair)	Maastricht University
Prof. dr. P. Santtila	NYU Shanghai
Prof. dr. J.A.A.C. Claessen	Maastricht University
Prof. dr. P.-A. Granhag	University of Gothenburg
Dr. M.J. Dubelaar	Radboud Universiteit

This research is supported by a fellowship awarded from the Erasmus Mundus Joint Doctorate Program The House of Legal Psychology (EMJD-LP) with Framework Partnership Agreement (FPA) 2013-0036 and Specific Grant Agreement (SGA) 532473-EM-5-2017-1-NL-ERA MUNDUS-EPJD to Enide Maegherman.

Table of Contents:

Chapter 1	Theories on Legal Decision-Making and Falsification	7
Chapter 2	Evidence Consideration in Practice: A Case Study	33
Chapter 3	Judges' Views, Experience and Knowledge of Falsification: An Exploratory Survey	59
Chapter 4	Law and Order Effects: On Cognitive Dissonance and Belief Perseverance	87
Chapter 5	Cognitive bias and Accountability in Legal Decision-Making	115
Chapter 6	Test of the Analysis of Competing Hypotheses in Legal Decision-Making	141
Chapter 7	General Discussion	163
	References	182
	Impact Paragraph	192
	Summaries	195
	Acknowledgements	207
	Research Dissemination	211
	Curriculum Vitae	213
	Appendices	214



CHAPTER 1

THEORIES ON LEGAL DECISION-MAKING AND FALSIFICATION

This chapter is an extended and more theoretically developed version of the publication: Maegherman, E., Horselenberg, R., Ask, K., & Van Koppen, P. J. (2019). From pervasive beliefs to wrongful convictions, *The Inquisitive Mind, 1/2019.*

1.1. When Decisions Go Wrong: The Schiedam Park Murder Case

On the 22nd of June, 2000, a ten-year old girl, Nienke was murdered in a park in Schiedam, the Netherlands. Her friend, Maikel, who was with her at the time of the attack, was also assaulted. Despite being stabbed, Maikel managed to survive, and could give a description of the perpetrator. Several witnesses were interviewed, including Cees B., who would later become a suspect. In July 2000, a policeman's son recognised Cees B. as the man who had proposed sexual acts to him the previous year. The policeman then confronted Cees B., who admitted what he had done the year before and said he wanted to get help to cope with the paraphilia. They made an appointment for Cees B. to come to the police station for a conversation. When the policeman looked up Cees B. in the police system, he found out that Cees B. was also mentioned in the investigation of Nienke's death. He then contacted the team responsible for that investigation and told them about the incident involving his son and Cees B. The team became interested in Cees B. In his report of the investigation, Posthumus (2005) describes that a lot of the investigation was focused on Cees B. prior to his arrest on the 5th of September, 2000, but that that was not the only focus of the investigation. Cees B. confessed on the 9th and 10th of September, after which the investigation team was reduced in size. From the 11th of September onwards, Cees B. retracted his confession and continued to deny he had attacked the children (Van Koppen, 2003). Cees B. was convicted of the attack on the children. A few years later, another man, Wik H., confessed to attacking Nienke and Maikel. An investigation resulted in incriminating evidence against Wik H., and he was convicted of the attack on Nienke and Maikel. It became clear that Cees B. had been wrongfully convicted, and an evaluation of the investigation was ordered. According to Posthumus' evaluation (2005), the investigative measures after the confession had almost exclusively been focused on Cees B and the confessions were vital for the evidence against him.

Cees B. confessed to the attack, but later retracted his confession. There were several indications that his confession was false, even before Wik H. confessed to

the crime (Van Koppen, 2003). Due to the interviewing techniques that were used during the interrogations with Cees B., as well as the inconsistencies between his confession and the witness statement made by Maikel, the confession should not have been believed (Van Koppen, 2003; Van Koppen, 2008). Instead, the police attempted to resolve the discrepancies by interviewing Maikel. When he did not change his statement, expert reports were used to conclude that the intense emotions of the event affected Maikel's perception and that his statements could not be trusted (Van Koppen, 2008).

Although it has been argued that the conviction in the Schiedam Park murder case was not a miscarriage of justice because there was sufficient evidence against Cees B., for instance by several judges in a documentary on judicial decisions (Verbraak et al., 2015), it has also become clear that there was exonerating evidence that was ignored or insufficiently appreciated in this case. Besides the retraction of the confession, there was also a discrepancy between the description of the perpetrator given by Maikel, and DNA of an unknown individual that did not match Cees B.'s profile.

DNA was found underneath Nienke's nails, on her stomach, on her left boot, and on the shoelace around her neck. The DNA that was found contained profiles of Nienke, Maikel, and of a third unknown individual. There was disagreement between the researcher and the expert of the Dutch Forensic Institute about the interpretation of complicated mixed profiles of the DNA on the lace and her stomach. According to the researcher, the markers all pointed at the same unknown male. However, the expert considered those profiles to be insufficiently reliable. They were therefore classified as no profile and were left out of the final report. The possibility that the unknown profiles came from the same male individual could, however, not be excluded. The Dutch Forensic Institute concluded that the DNA profile of the unknown individual could not be related to the profile of Cees B (Posthumus, 2005). That evidence could have undermined the confession given by Cees B., or should have given rise to alternative routes of investigation.

In addition to someone else's DNA being found on the victim, there was also the description of the perpetrator given by Maikel, the surviving victim of the attack. The description he gave was very detailed. He described a white male, with a normal posture and a pale appearance. He was described as having red and white spots in his face, several of which had been scratched open. The police initially tried to find the person matching the description. Cees B. did not fit the description given by Maikel (Van Koppen, 2003). The discrepancies between Maikel's statement and Cees B.'s appearance and confession were not addressed by the investigation team and the confession was used by the court (Posthumus, 2005).

1.2. Aims of the Current Thesis

The Schiedam Park murder case illustrates one problem that has contributed to several miscarriages of justice, namely the excessive focus on incriminating evidence and the lack of appreciation for exonerating evidence. Upon consideration of several of the cases listed by the Innocence Project (Innocence Project, 2020), it becomes clear that that process plays a role in many miscarriages of justice, although it may not always be acknowledged. The consideration of exonerating evidence in legal decision-making is the focus of the current thesis. More specifically, in the current thesis, I investigate how judges use falsification, or consider exonerating evidence, and how it could be facilitated through changes to the existing procedure, such as the construction of the case file or the instructions to explain the decision on guilt.

The current thesis serves two aims. The first aim was to determine the role of falsification in practice in the context of legal decision-making. In order to determine whether judges are firstly, aware of the concept of falsification, and secondly, willing and able to attempt falsification in practice. A case study and a survey were conducted in order to elucidate how the current practice relates to the theoretical understanding and approaches. The second aim of the thesis was to research potential ways in which falsification could be facilitated in

the decision-making. The various methods which were tested were based on other areas of research, where positive effects have been found in reducing the focus on an initial belief of guilt, and thereby increasing the use of exonerating evidence in decision-making.

1.3. Legal Decision-Making

In order to understand how falsification plays a role in legal decision-making, it is firstly important to consider legal decision-making in general. For the purpose of this thesis, legal decision-makers refers to those who are given the task to decide whether or not a defendant is guilty of what they are accused of. In other words, the focus of this thesis is on the decision of guilt, rather than on decisions leading to prosecution or on decisions of sentencing. Depending on the jurisdiction, a decision on guilt can be made by either one or several judges, or by a jury of lay people. As the decision-makers never witness whether the defendant had the alleged role in the event that took place as described in the charge, the decision-makers inevitably have to bridge an inherent gap between the evidence and what they are convinced has happened. Regardless of whether the decision is made by a jury or a judge, the available evidence still needs to be weighed in order to come to a decision. Although countries usually have legal prescriptions as to what can be used as evidence and what cannot, these regulations tend to offer little help when it comes to how convincing a piece of evidence is, or what the weight of that particular piece of evidence should be (Anderson et al., 2005).

When deciding whether a defendant is guilty of what they are accused of, a judge needs to be convinced that the offence happened as described by the prosecution. In the Netherlands, the evidence used for that decision needs to be legal and convincing (Verbaan, 2016). Other countries describe the conviction threshold as being convinced beyond a reasonable doubt. Similarly to the lack of research shedding light on how evidence is weighed, little is also known about how the threshold of beyond a reasonable doubt is reached. Although it

is usually clear whether the minimum legal requirement for evidence has been met, whether or not that evidence is also convincing is an entirely different question (Wagenaar et al., 1993). If the court is not convinced of the suspect's guilt based on the evidence, the suspect must be acquitted. Similarly, if the court is convinced but there is not sufficient legal evidence, the suspect must be acquitted (Dubelaar, 2014).

1.3.1. Theories on Legal Decision-Making

Several theories have been proposed to explain how evidence is weighed in order to come to a final decision on guilt. Wigmore was among the first to theorise on the concept of proof, namely the process of persuasion rather than on the procedural rules of admissibility (Twining, 1985). Wigmore's theory was largely based on inductive logic, and he believed that the reasoning with regards to judicial proof should be in line with the reasoning and decision-making in which people engage in everyday life. He considered facts to be propositions, and evidence as the relation between a proposition to be proved and the proposition which supported it. A proposition supporting another proposition could in turn be supported by yet another proposition, thereby forming a chain, which is common in judicial trials. At each step, doubts may be raised, and the propositions can be challenged. That could, for instance, happen in the form of an alternative explanation for the facts. The alternative explanation would then reduce the value of the conclusion that can be drawn from the information, namely the inference. As Bex (2010) explains, that reduced value can be due to the generalisation which accounts for the inference not being applicable in the particular context, thereby invalidating the inference between propositions rather than the proposition itself. It could also happen by negating the facts or propositions, such as for instance using an alibi to negate the assumed fact that the suspect was at the scene of the crime (Twining, 1985). The different chains and challenges would result in a chart, which represents the logical relations

according to the individual who makes the chart. The probative value of evidence is limited to the conclusion being the result of the most likely inference.

For example, if the main fact to be proven is that the defendant shot X, then the propositions which support that fact could include that the defendant carried a gun, that he was at the crime scene at the time of the shooting, and that a witness claims he saw the defendant shoot X. Doubt could then be raised by a proposition that the witness is lying about what he saw. The inference following from the witness statement, namely that the defendant shot X, may then no longer be applicable. Furthermore, if the defendant provides an alibi for the time of the murder, the proposition that he was at the crime scene is also negated. Although the example is here intended to be simple, when using Wigmorean charts in real-life cases, the result is often complex (e.g., McDermott, 2015).

As will become clear through the explanation of later theories, several elements of Wigmore's theory provided a basis for the further development of theories on legal decision-making. Within theories on legal decision-making, two schools can be distinguished: the argument-based approach and the story-based approach (Bex et al., 2010). According to Bex et al. (2010), the psychological perspective focuses on using stories for evidence organisation and analysis. As the theory of the current thesis is also mostly related to the story-based approach, some of those theories will here be described in more detail.

Bennett and Feldman (1981/2014) describe several explanations for the use of the story structure within the context of legal decision-making. They argue that there is a need for a framework that can be understood by both those who received legal training and those who have not, such as jurors, but also witnesses or defendants. According to Bennet and Feldman (1981/2014), the formality of the legal procedures needs to correspond to informal, commonplace social judgments made by individuals in such a way that the relevant legal questions can be understood in everyday terms. Using a story can help with that understanding as it is a common form of communication about complicated events. For instance, Adaval and Wyer (1998) found that information presented

to consumers was better processed when given in the narrative form compared to when it was presented as a list. Furthermore, the story format can help in the interpretation of the central elements of the case. The interpretation of evidence using the format of a story was further developed by Pennington and Hastie in their Story Model (1986) of legal decision-making.

1.3.1.1. The Story Model. One of the earliest versions of a story-based approach, primarily developed and tested in the context of juror decision-making, was the Story Model presented by Pennington and Hastie (1986). According to the story model, evidence is evaluated through the creation of a story, and the decision is then made by seeing which verdict category the story best fits into. Pennington and Hastie (1986) argue that the organisation of evidence helps jurors to better understand the evidence. The story is constructed during the presentation of evidence at trial, as well as after. By incorporating the evidence into the story, it also becomes meaningful. The organisation of the story is based on relations which imply cause and intentions of actions, according to the jurors' general knowledge concerning action sequences (Pennington & Hastie, 1986). The story that is constructed uses the information about the case given during the trial, as well as knowledge about events that are similar to those in question, and lastly, expectations of what is needed for a story to be complete (Pennington & Hastie, 1992). In addition to constructing a causal model to explain the information available to the decision-maker, subsequent decisions are also thought to be based on the causal explanation that the decision-maker has constructed (Pennington & Hastie, 1988). Although several stories may be constructed, one will be preferable to the others. Pennington and Hastie (1992) explained the preference in terms of how well the story can account for the evidence, as well as how coherent the story is. Coherence was explained as a combination of consistency, completeness, and plausibility. Pennington and Hastie (1988) define consistency as the lack of contradictions, and plausibility can be determined by comparing the story to other events in the real world. Moreover, by adding causal story connections to case materials, the evaluation

of evidence also changed in line with the story direction (Pennington & Hastie, 1992).

The next phase of decision-making according to the story model involves the juror becoming informed about the decision alternatives. Each of the alternatives has specific defining features. Pennington and Hastie (1991) emphasized that the majority of criminal cases have several verdict alternatives, defined by a number of elements. These can be related to the different elements of the charge which must be proven, such as for instance the intention of the perpetrator or the offence to which the actions correspond. For example, if someone is killed, the perpetrator might be convicted for manslaughter or murder, for which different elements (e.g., premeditation) would have to be proven. Another verdict category could be that the perpetrator acted out of self-defence, and is considered not guilty. In the final phase of the story model, the features of the story constructed by the juror are matched to the elements of the different verdict categories in order to come to a decision (Pennington & Hastie, 1992).

The story model was also tested by Pennington and Hastie. For instance, they found that the party who presented their evidence in a story format was judged more favourably than when evidence was organised by legal issue (1992). Similarly, they found that jurors remembered evidence that was congruent with their story better than evidence that was not (1988). Subsequent theories, that also widened the scope outside juror decision-making, were based on the story concept akin to what was described by Pennington and Hastie (1988; 1992; 1993). To give an example of how evidence can be incorporated into a story, the case of the defendant accused of killing X can again be used. The elements that can make up a story might include for instance that the defendant previously had problems with X. The inference based on other knowledge may then be that the defendant had a motive to kill X. That psychological condition can then be combined with case specific information, such a witness statement stating that the defendant took a gun to his meeting with X. From the physical condition of

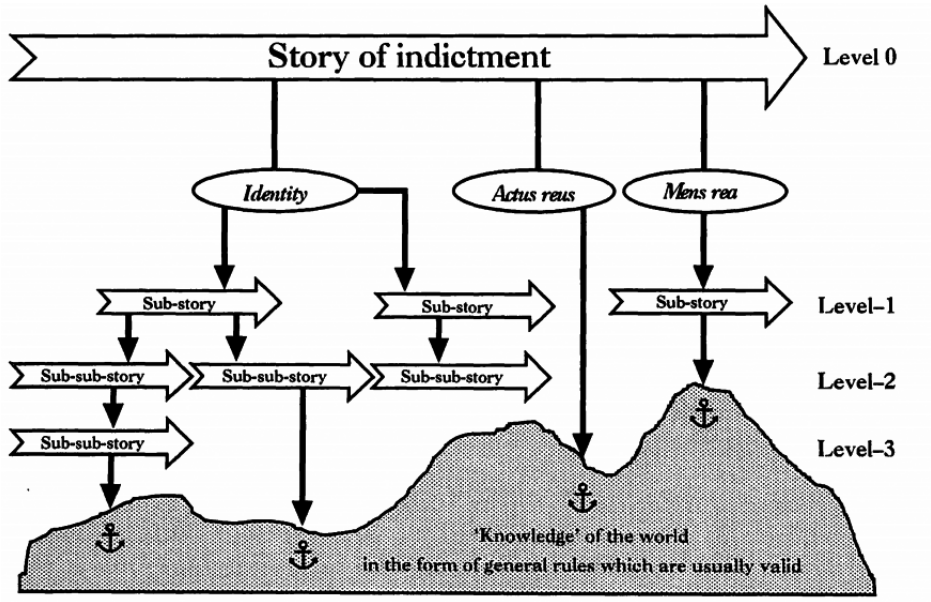
taking a gun, the inferred goal could then be that the defendant intended to kill X and shot him, which is the action.

1.3.1.2. Anchored Narratives. The theory of anchored narratives can be considered a further development of the story model. Wagenaar and colleagues (1993) describe the decision made on guilt as based on two judgements, namely the plausibility of the prosecution's version of events, and whether that version can be anchored to common-sense beliefs by the evidence. In addition to what constitutes a good story, as described by Pennington and Hastie (1992), the theory of anchored narratives also incorporates the need for the story to be true, as established by the evidence (Wagenaar et al., 1993). In order for evidence to be able to prove something, it requires the belief in a general rule that can be considered valid most of the time. Wagenaar et al. (1993) give the example that a witness statement can only prove something if it is believed that witnesses do not lie and do not make mistakes. As there are exceptions to rules, it may be required to prove that an exception is not applicable, for instance by supporting the witness statement with other evidence. Every piece of evidence needs to be anchored into a generally accepted rule (Wagenaar et al., 1993).

The theory of anchored narratives is illustrated in Figure 1, which shows that the different elements of the indictment, and their substories, need to be anchored into commonly accepted knowledge. The story of the indictment forms the main story to be anchored. The evidence is ordered so that it acts as an anchor of the story in rules that are commonly accepted. The evidence itself also forms a substory which in turn needs to be anchored by additional evidence. That additional evidence then becomes a sub-sub-story and again requires anchoring. A sub-story can be anchored when the common sense rule it represents is accepted as true (Wagenaar et al., 1993).

Figure 1

Schematic Overview of the Theory of Anchored Narratives (Wagenaar et al., 1993)



In order for the story of the indictment to be anchored, the structure does not need all evidence to be included. The evidence included is only used to verify the indictment. There is no need for evidence that contradicts the indictment (i.e., falsifying evidence) to be included in the structure when anchoring the indictment (Wagenaar et al., 1993). Therefore, when using the structure of anchored narratives to come to a decision, not all the evidence needs to be considered in order to be convinced of the indictment.

The theory of anchored narratives can be explained using an example. If the defendant is accused of murdering X by shooting him in the head, then the shooting of X is the central element. It then needs to be anchored that the defendant is the one who shot X (identity), that X was unlawfully killed (*actus reus*), and that the defendant had the intention, and premeditation, to kill X (*mens rea*). These elements then need to be anchored. For instance, the intention could be divided into the sub-story that the defendant had told a police informant

he was going to kill X, and the fact that he brought a gun to his meeting with X. The judge could at that point anchor the substories in the general rules that police informants are reliable witnesses, and that when taking a gun to a meeting, there is an intention to cause harm. Of course, these substories could be further divided into other substories, but for the simplicity of the example, the substories are in this case anchored quickly. It is up to the court to determine when the story is sufficiently safely anchored, although there is no such thing as absolutely safe anchors (Wagenaar et al., 1993). Unsafe anchoring would for example happen if the only possible anchor is one for which a large number of exceptions is possible, or if an anchor is based on an incorrect belief (e.g., that suspects do not falsely confess).

Unsafe anchors can result in unsafe convictions. As various evidence differs in what is commonly accepted for it, the amount of substories required to anchor the evidence also differ. Furthermore, Ask and colleagues (2008) suggested that the rules used for anchoring can be affected by bias. Participants in their study rated the reliability of eyewitness evidence in general differently, depending on whether or not an evidence statement in the specific case was consistent with their initial belief about the case. Therefore, the rules that are used for anchoring the evidence may not be stable across cases, which can also be related to the lack of consistency in determining the weight of evidence.

Similarly, Wagenaar et al. (1993) predicted that, according to the theory of anchored narratives, judges' selection of evidence would be biased, as the focus is on anchoring the main narrative, namely the indictment. That is also in line with the requirement to explain a guilty decision according to the Dutch Code of Criminal Procedure, which requires the judge to give reasons and evidence to support the conviction (Art. 359 DCCP). Although the code requires an explanation as to why explicit arguments made by the defence or the prosecution were deviated from, there is no requirement for all the evidence to be considered in judges' decisions. Thus, it can be expected that, if an indictment can be sufficiently anchored, evidence that contradicts the indictment may be

unlikely to affect the support for the indictment, and the decision to consider it proven.

In their book on anchored narratives, Wagenaar et al. (1993) argued that the heuristic described by the theory of anchored narratives is relevant to both judges and juries, and unlikely to differ across cultures. Viewing the anchored narratives theory as a model of general human reasoning can also be interpreted as a development of Pennington and Hastie's story model (1986), which focused mainly on juror decision-making. Wagenaar et al. (1993) offered ten rules of evidence that they considered to be universal. In one of these rules it is stated that there should be no other story which has equally good, or better, anchoring. That rule provides a key basis for the scenario approach.

1.3.1.3. The Scenario Approach. The scenario approach, as explained by Van Koppen and Mackor (2019), is closely related to the theory of anchored narratives, although more focus is placed on the consideration of multiple scenarios. A scenario offers a chronological and causal account of an event, including a central action that is understood in the context of the surrounding scene. For example, in the example above, the shooting is the central action which is understood in the context of X dying. As previously described by Pennington and Hastie (1992) in their description of a story, a scenario here also relies on background knowledge, or scripts.

The next component of the scenario approach is the inference to the best explanation. Harman (1965) describes the key inference to the best explanation as concluding, from the fact that a hypothesis is able to explain the evidence, that the hypothesis is true. Alternative hypotheses which could explain the evidence have to be rejected before making that inference. If one particular hypothesis can give a better explanation for the evidence than any other alternative hypothesis, it can be inferred to be true (Harman, 1965). Therefore, a comparison of the different scenarios in terms of how well they explain the evidence is a key element of the scenario approach. Pardo and Allen (2008) argue that the outcome

of the inference to the best explanation is not guaranteed to be the truth. The selection of the alternative scenarios is therefore key in the scenario approach.

Van Koppen and Mackor (2019) provide several criteria for selecting and assessing scenarios, divided into internal coherence of scenarios, the coherence of the scenarios with general background knowledge, and coherence of the scenarios with elements of the case that are accepted as true. The second and third requirements relate back to the explanation of the theory of anchored narratives described above, where the sub-scenarios of the main scenario are ultimately anchored in commonly accepted knowledge. Van Koppen and Mackor (2019) explain three ways in which scenarios are related to evidence: creation, accommodation, and prediction. A scenario is *created* during a criminal investigation on the basis of evidence. If evidence is encountered that is not in line with the scenario, the conclusion could be that the scenario should be rejected, or the scenario should be adapted in order to *accommodate* for the evidence. The scenario can also be used to *predict* evidence one would expect to find.

In addition to the coherence of scenarios, Van Koppen and Mackor (2019) also emphasise the need to try to falsify the scenario. Falsification was considered a key step in reasoning by Popper (1959/2005). He argued that in addition to looking for evidence that can be explained by a scenario, it is important to look for evidence that can disconfirm, or falsify, a scenario. A theory or scenario can only be considered true until evidence that disconfirms it is found. If one only tries to find evidence supporting one's theory, then it may be incorrectly considered as true. On the other hand, if one actively tries to find evidence against one's theory, but fails to do so, it becomes more likely that the theory is true, as no evidence disproving it can be found. Attempts to falsify a scenario are equally as important as attempts to verify a scenario. Specifically, failed attempts at falsification increase the likelihood of the scenario being true (Crombag et al., 2006).

Van Koppen and Mackor (2019) propose that finding a good alternative scenario can also be considered part of the falsification process. While the indictment is based on a coherent selection of evidence by the prosecution for the scenario that the suspect is guilty, it is possible for the defence to construct an alternative scenario based on another selection or interpretation of the evidence in the case file showing that the suspect is not guilty, or that the events did not happen in the way suggested by the prosecution. Those scenarios should then be compared to one another, although that is not commonly done by judges (Crombag & Wagenaar, 2000). Evidence will not only support one possible scenario, but rather, it can discriminate between scenarios. Stronger evidence will be better at discriminating between alternative scenarios (Van Koppen, 2011). It has been suggested that the explanation requirement imposed on Dutch judges, according to which a reasonable alternative scenario proposed by the defence needs to be excluded, does not need a comparison of two, or multiple, scenarios to determine the value of the evidence (Stevens, 2017).

1.4. Confirmation Bias

Although falsification and the consideration of alternative scenarios are vital steps in determining whether or not a scenario is true, it appears to be obstructed by the inherent problem in human reasoning known as confirmation bias. Confirmation bias can be defined as a tendency to favour evidence that supports an existing belief in both the seeking and interpretation of evidence, while paying disproportionally little attention to evidence that contradicts the belief (Kassin et al., 2013; Mendel et al., 2011; Nickerson, 1998). Confirmation bias has been considered a contributing factor to miscarriages of justice (Rassin, 2010). It is thought to be particularly relevant to criminal law proceedings (Findley & Scott, 2006)

Several theories explaining the underlying mechanism of confirmation bias exist. Throughout this thesis, the mechanism that has been adopted and tested is one that considers confirmation bias to be related to cognitive dissonance and

belief perseverance. Cognitive dissonance can be described as an uneasy feeling that one experiences when confronted with conflicting information. The term was introduced by Festinger (1957), who proposed that one who experiences dissonance will try to reduce it by achieving consonance. He argued that the existence of cognitions that are incongruent can be a motivating factor due to the psychological discomfort created. He further explained that cognitive dissonance could be seen as a precursory condition to behaviour aimed at reducing dissonance.

The way in which cognitive dissonance is experienced according to Festinger's theory (1957) was tested empirically by Elliot and Devine (1994), who tested participants' discomfort when writing an essay that either contradicted or supported their own attitudes. They found that those who agreed to write a counter-attitudinal essay experienced more discomfort, which returned to a baseline level when participants had changed their attitudes in line with the essay. Their findings support the fact that dissonance is experienced as a psychological discomfort. They also found support for participants being motivated to reduce dissonance. Harmon-Jones and Harmon-Jones (2007) conducted a review of cognitive dissonance literature since the conception of the theory and concluded there is more support for the original theory as proposed by Festinger than for alternative explanations of dissonance, such as that it is due to an expectation of negative consequences.

Festinger (1957) states that the presence of dissonance will influence the subsequent information seeking and selection. Cancino Montecinos (2020) outlined several dissonance reduction processes in an overview of earlier research. These include *attitude change*, which involves changing of the negative attitude to become more positive. *Trivialisation*, which appears to be closely related to the manifestation of confirmation bias, involves reducing the importance of the original attitude or of the dissonant information. For example, smokers are more likely to refuse information that smoking is causally related to lung cancer (Festinger, 1957). Similarly, *bolstering* of the initial attitude in

response to a contradicting attitude can also be related to favouring information confirming an initial attitude (Cancino Montecinos, 2020).

Cognitive dissonance, or rather consonance, can also be conceptually related to the cognitive coherence view of legal decision-making that was proposed by Simon (2004). Simon (2004) suggests that coherence is imposed on cognitively complex tasks, such as legal decision-making, in order to transform it into a simpler one, so that an easier and confident decision can be made. Simon et al. (2004) also demonstrated that the process of achieving consonance or consistency between the evidence and a legal decision is bidirectional: the evaluation of the evidence influences the decision, but a preference for a decision also influences the evaluation of evidence. The need for coherence is also relevant to the story-based theories on legal decision-making described above, as the need for evidence to be in line with a story could influence how evidence is interpreted or whether it is selected by the decision-maker.

In order to achieve consonance, one may show belief perseverance, the tendency to ignore evidence or information that is contradictory to one's belief. Belief perseverance can therefore help in reducing the uneasy feeling associated with dissonance as has been experimentally demonstrated by Green and Donahue (2011). They presented people with a story, and then told them that the story was false. Regardless of whether participants thought the initial information was false due to intentional deception or accidental error, their story-related beliefs remained unaltered. Similarly, belief perseverance has also been observed in practice in the context of wrongful convictions. Some prosecutors in cases where the suspect was wrongfully convicted and later exonerated maintain their initial belief in the guilt of the suspect (Burke, 2007).

Belief perseverance is conceptually related to confirmation bias. Confirmation bias extends the manifestation of belief perseverance from ignoring contradicting evidence to excessively focusing on confirming evidence. Wason (1960; Wason & Johnson-Laird, 1972) conducted some of the first studies in which confirmation bias was demonstrated, although the term had not yet been introduced. Wason

conducted two classical studies which show the human tendency to try and confirm a hypothesis or a rule. In one of Wason's studies (1960), participants were asked to determine a rule. The only information people were given was that the sequence 2-4-6 fit the rule. They were then given the opportunity to give additional sets of numbers, and would be told whether or not that set of numbers fits the rule. They were also told that they should use as few sets of numbers as possible to test the rule they had in mind. They should only state the rule when they were very confident. Participants showed a tendency to try and confirm their rule by listing three numbers that fit the rule they believed to be correct. For instance, if they thought the rule was 'increasing by 2', then they would suggest 8-10-12. Those who eliminated more possibilities and tried more incorrect sets of numbers were more likely to give the correct rule on the first attempt at stating the rule.

Another famous study in which Wason showed the existence of confirmation bias was done using a card selection task (Wason & Johnson-Laird, 1972). Participants were given the conditional rule that if there was a vowel on one side of the card, then there was an even number on the other side of the card. They were given cards with the following symbols on one side: E, K, 4, and 7. They were given the task to name those and only those cards that had to be turned over in order to determine whether the rule was true. Most participants selected either only the E or the E and the 4 card, thereby only trying to confirm the rule. The correct answer would be to turn around the E and the 7 card, as an odd number on the E or a vowel on the other side of the 7 would falsify the rule. Wason and Johnson-Laird (1972) emphasised that participants in the card selection task failed to select the card that could falsify the rule, and selected a card which could not falsify the rule. The Wason selection task has been used in various formats to test for confirmation bias since it was first introduced (e.g., Jones & Sugden, 2001; Rachlinski, 2012; Rassin, 2018).

Confirmation bias has been shown in a variety of legal-psychological contexts. Dror et al. (2006) demonstrated the effect of biasing information on

the interpretation of fingerprint evidence. Several studies have demonstrated the effect of a prior belief of guilt on interviewing as well as the subsequent perception of those interviews by neutral observers (Kassin et al., 2003). In addition to confirmation bias affecting several stages of a criminal investigation, it can also affect the interpretation, evaluation, and integration of evidence.

1.4.1. Confirmation Bias in Reasoning with Evidence

Little research has been conducted in practice on the topic of how judges reason with evidence. One study was conducted by Schünemann (1983, as cited in Schünemann & Bandilla, 1989). They compared a group of judges who had received mainly incriminating information prior to trial to a group of judges who had received information that was less incriminating. The groups read identical trial proceedings. Of the former group, 82% would have convicted the suspect, compared to 53% in the latter group. Based on Schünemann and Bandilla's study, it seems that judges can also be affected by an initial belief in such a way that prior beliefs affect the eventual guilt decision.

In several experimental studies, support has been found for the prediction that confirmation bias also affects the reasoning with evidence. For instance, Eerland and Rassin (2012) presented participants with a case file with exonerating or incriminating investigative measures that had either been successful or failed. They found that participants who already believed in the suspect's guilt became more convinced in response to incriminating evidence, but were reluctant to change their impression of guilt in response to exonerating evidence. Moreover, successful attempts to find evidence (whether incriminating or exonerating) had more of an impact on guilt perceptions than failed attempts to find evidence (e.g., finding fingerprints increased the perception of guilt, but not finding them did not decrease the perception of guilt). Rassin et al. (2010) found that law students' selection of investigative measures was influenced by their initial assessment of guilt. They concluded so based on three studies in which participants had to first read a case file and then had to select either incriminating or exonerating

investigations. Those who initially thought that the suspect was guilty favoured incriminating investigations, whereas those who thought the suspect was innocent did not.

Similar findings were also reported by O'Brien (2009), who found that participants who expressed a hypothesis about a prime suspect were biased towards incriminating evidence against the suspect while reading the remainder of the case file. Ask and Granhag (2005) found evidence for confirmation bias, in that investigators' interpretation of the same set of evidence could be influenced by presenting them with one of two alternative initial hypotheses. Ask et al. (2008) found that participants were more sceptical towards evidence that was inconsistent with their prior belief than to evidence that was consistent. In line with the findings by Ask et al. (2008), a later study by Ask et al. (2011) found that this asymmetrical scepticism towards evidence also differed between different types of evidence, which the authors termed the 'elasticity' of the evidence. For instance, eyewitness testimony is more elastic than DNA evidence in that the interpretation of the evidence can be more subjective. Overall, the results of a large collection of studies strongly support the proposition that reasoning with evidence in the context of criminal law proceedings can be biased by a decision maker's initial belief.

1.5. Debiasing Techniques

Considering the potential detrimental impact that biases can have on reasoning, there is a keen interest in trying to reduce bias and its consequences. Several researchers have found support for the bias blind spot, which means people can be aware of a bias in others without thinking it affects themselves. Kukucka et al. (2019) recently found support for a bias blind spot in forensic examiners. Therefore, mere awareness of the existence of bias appears to be inefficient in debiasing people. Fortunately, other debiasing strategies have proven to be more effective. For instance, in the medical field, the use of differential diagnosis, whereby a decision is made between several possible

conditions, has been offered as a way to reduce the intuitive biased judgement. Shimizu et al. (2013) found that differential diagnosis resulted in better diagnostic performance in difficult cases, but not in simple cases. Such applied approaches may therefore be useful as a debiasing strategy.

In addition to using procedures to counter bias, several attempts have also been made at developing training which can successfully counter bias. Morewedge et al. (2015) developed a computer game and an instructional video to reduce biases in the context of intelligence analysis. The game involved teaching people about bias and the influence of the bias, and providing simple and extended feedback, the latter of which included strategies for coaching, intervention and mitigation. They found the interventions to reduce bias, even when the participants were tested several months after the intervention. Shaw and colleagues (2018) also mitigated bias in participants using an educational game, with a mitigating effect being observed eight weeks after the intervention.

In the context of legal decision-making, relatively little research has been conducted on how confirmation bias can be reduced, or how falsification attempts can be increased, despite several researchers finding support for the effects of bias in that context (O'Brien, 2009; Rassin, 2018; Rassin et al., 2010; Ask et al., 2008, 2011). Little research has focused directly on judges' consideration of exonerating evidence. However, several studies have investigated the use of alternative scenarios. For instance, O'Brien (2009) found that considering evidence for why an initial hypothesis might be wrong reduced confirmation bias. However, considering evidence for multiple scenarios did not reduce bias, contrary to expectations. Rassin (2018) used a pen and paper tool to aid participants in their consideration of alternative scenarios. Participants had to rate how well each piece of evidence fit with a primary or an alternative scenario. Rassin (2018) found that participants using the tool showed less confirmation bias than those who were simply presented with an alternative scenario. Tenney and colleagues (2009) found that when defence lawyers in a mock case argued for an alternative scenario in which another individual could have committed

the crime, in addition to refuting the incriminating evidence, participants returned fewer guilty verdicts than when the defence lawyer only rebutted the incriminating evidence. Based on these studies, it appears that the consideration of alternative scenarios has the potential to reduce the influence of bias, although the application is not as simple as might be expected. In a recent study, Sauerland and colleagues (2020) also found support for the difficulty in using alternative scenarios to counter bias in the context of allegiance bias, whereby experts are influenced by the party that hired them. As the instruction to consider evidence for and against three given scenarios did not reduce allegiance bias, it should be researched whether and how alternative scenarios can be used in such a way that the scenario that best explains the evidence becomes the chosen alternative, rather than the scenario which has the most supporting evidence. A scenario that has a lot of support may not be the scenario that best explains the evidence if there is a lot of evidence that does not fit with the scenario. Thus, both the use of alternative scenarios, as well as how to increase the consideration of exonerating evidence, requires further research to be conducted.

1.6. The Current Thesis

Miscarriages of justice have grave consequences for the wrongly convicted as well as for society as a whole, as they can undermine trust in the criminal justice system and cause false beliefs about crime in general (Cole, 2009). Miscarriages of justice are often due to a combination of a false belief in guilt, possibly due to problematic evidence such as a false confession or wrongful identification, and exonerating evidence being overlooked.

The aims and research of the current thesis can be divided into a practical and an experimental approach. In the former, the aim was to determine the role of falsification in practice in the context of legal decision-making. Specifically, to determine whether judges are firstly, aware of the concept of falsification, and secondly, willing and able to attempt falsification in practice. As discussed above, confirmation bias is considered to be especially important in the context

of legal proceedings. In the Netherlands, there are additional factors which I suspected may aid the influence of confirmation bias. These include for example the construction of the case file, which consists mainly of evidence that is incriminating for the suspect (Crombag, 2017). Furthermore, there is also the requirement for judges to explain their decision to find the suspect guilty, which requires the judge to provide evidence the decision could rest upon (Art 359 DCCP). In case of a confession, the other evidence used for the decision requires less explanation. The experimental approach related to those factors, and was used to investigate potential ways in which falsification could be facilitated through procedural changes or training.

1.6.1. Overview of Chapters

The first two studies described in the current thesis focused on the use of falsification in practice. The first one is a case study (Chapter 2). In the case study, the decision and reasoning of the court were critically examined using the evidence in the case file. The diagnostic value of the evidence used by the court was discussed, as well as that of evidence that was not mentioned in the reasoned decision. The study made use of the theories that were described above, particularly the scenario approach.

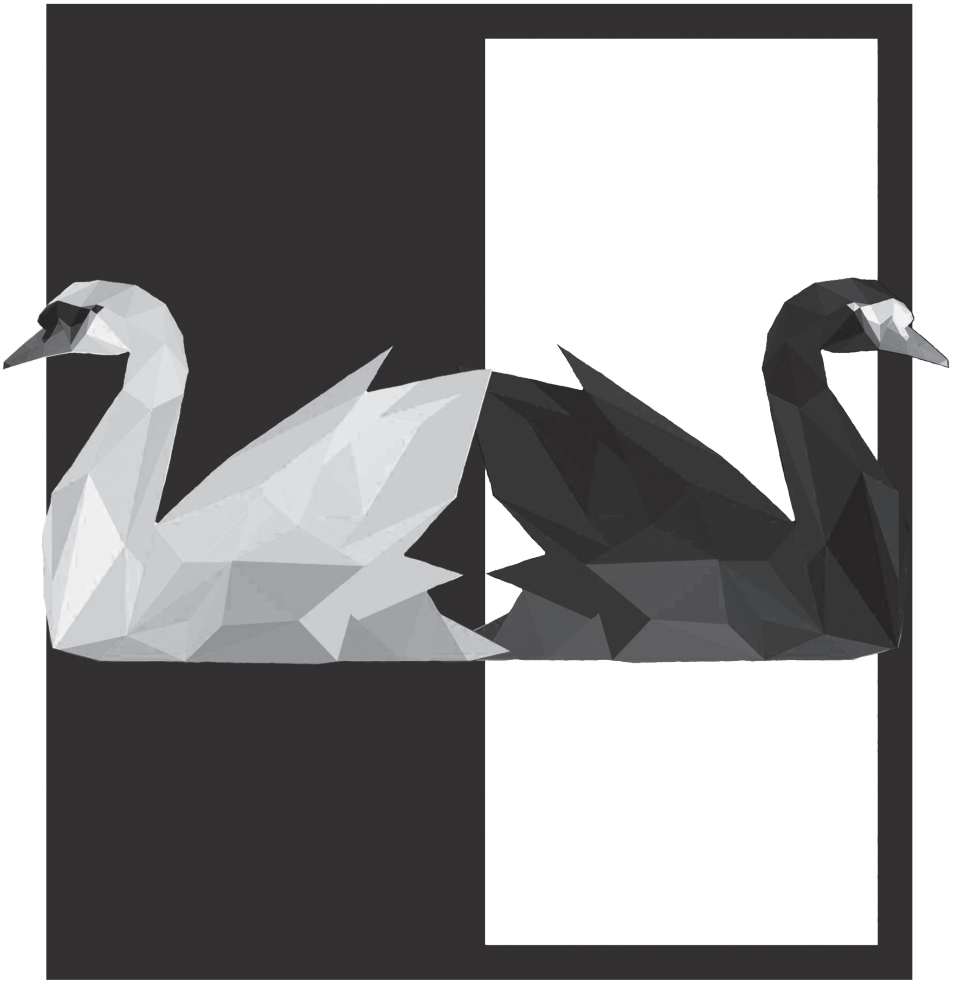
The second study, that I conducted to investigate the use of falsification in practice, was a survey conducted with judges (Chapter 3). The aim of the survey was to gain insight into judges' understanding of the concept of falsification, as well as how important they perceived it to be. The survey extended findings from a previous study also conducted with criminal law judges, in which the majority of respondents had expressed finding it important to consider alternative scenarios, as well as to look for evidence that contradicts a theory (Van Veldhuizen et al., 2019). The current study examined judges' understanding of falsification, their ability to apply falsification on a number of simple tasks, and their use of falsification in practice.

The remainder of the studies conducted for the current line of research investigated ways in which falsification could be facilitated. The first experimental study examined whether the order of evidence presentation influences the consideration of exonerating evidence (Chapter 4). A cognitive dissonance measure was included in order to determine whether there was a relation between the state of dissonance and the interpretation of evidence. As the judge normally prepares for the trial by reading a mostly incriminating case file, the possible influence of incriminating initial information warrants investigation. Based on the theory of confirmation bias, it was hypothesised that receiving incriminating evidence first would create a preference for later incriminating evidence, whereas receiving exonerating evidence first would create a preference for subsequent exonerating evidence.

In the second experimental study, I examined factors that may increase the use of falsification attempts, with a focus on the requirement for judges to explain their decision (Chapter 5). It has been established that needing to explain a decision can improve the quality of a decision (Lerner & Tetlock, 2003; Tetlock, 1983). Therefore, in this study, it was tested whether different instructions to explain a decision on guilt could increase the consideration of exonerating evidence. Specifically, experimental conditions were created on the basis of the Dutch Code of Criminal Procedure, the German Code of Criminal Procedure, and the principle of falsification. The German condition required, for instance, an explanation of other possible versions of events (Mevis, 2019). As the Dutch explanation requirement mainly asks for evidence that can support finding the defendant guilty, participants in the German and falsification conditions were expected to be less biased and to incorporate exonerating evidence to a larger extent in their final decision on guilt.

The final study was conducted to test the effectiveness of the Analysis of Competing Hypotheses (ACH; Heuer, 1999) to counteract confirmation bias in a criminal law setting (Chapter 6). ACH is a method developed for the context of intelligence analysis in order to reduce the influence of confirmation bias.

According to ACH, a matrix should be constructed which allows each piece of available evidence to be compared against each possible hypothesis. The method is focused on falsification, as the hypothesis that is chosen should be the hypothesis with the least contradicting evidence. It was expected that participants who received information on the ACH method would be less likely than those who received general information on biases to perceive the suspect as guilty in a biased murder vignette.



CHAPTER 2

EVIDENCE CONSIDERATION IN PRACTICE: A CASE STUDY

This chapter is a translated and adapted version of the following publication:
Maegherman, E.F.L., De Boer, M. M., Horselenberg, R., & Van Koppen, P.J. (2020) *De onzichtbare steekpartij: Over strategische selectie van getuigenbewijs. [The invisible stabbing: About strategic selection of witness evidence.]* Boom Criminologie.

Abstract

In this case study, the application of theories on legal decision-making in practice was considered. The reasoning process explained by the court in the written decision was explored using the evidence in the case file to determine whether attempts at both confirmation and falsification were made. Furthermore, the evidence used by the court to find the defendant guilty was examined for its ability to discriminate between different scenarios. The case concerns the fatal stabbing of a young man following a fight during carnival celebrations. Nobody saw who stabbed the victim. The defendant's conviction was, to a large extent, based on two witnesses who claimed the defendant had confessed to stabbing the victim. In this case study, the validity of these statements, as well as the consideration of other evidence was examined. Based on the reasoned decision, it seems the court made an effort to consider alternative scenarios, although the use of falsification appears to have been limited.

2.1 Introduction

In this chapter the theory that has been discussed in the previous chapter has been applied to a case. The case has been the focus of a more extensive discussion, which can be found in the book *De onzichtbare steekpartij* (The invisible stabbing; Maegherman et al., 2020). The discussion in this chapter has been limited to the facts of the case that are relevant to the focus of this thesis, namely the consideration of evidence by the court. Although, at times, facts are discussed in detail, most details of the police investigation were not addressed in this chapter. The focus here is on the decisions made by the court, and the reasoning and considerations which led them to find the defendant guilty.

When deciding on a case, the court, often consisting of multiple judges, has the task to determine what happened based on the evidence provided to them by the police in the form of a dossier or case file. The court has to be convinced that the suspect committed the crime if it chooses to convict him. If the court is not convinced of the suspect's guilt based on the evidence, the court must acquit. Similarly, if the court is convinced but there is not sufficient legal evidence, the suspect must be acquitted (Dubelaar, 2014). Becoming convinced on the basis of the evidence requires an inherent leap, as the evidence can rarely inform the court with absolute certainty of what actually happened, and so inferences always need to be made. These inferences are at the core of this thesis. In most cases, a decision needs to be made based on multiple related or individual pieces of evidence, which range from very relevant to not at all relevant, as well as vary in diagnostic value. The evidence not only needs to be interpreted by the court, but also needs to be selected for relevance (Van Koppen & De Keijser, 2017). Ultimately, the evidence will also have to be integrated to come to a final decision on guilt.

In the Netherlands, the court is restricted in what evidence it can base its decision on. According to the Code of Criminal Procedure (Art. 339), the permitted evidence is limited to: witness statements, suspect statements, expert statements, the judge's own perception, and written documents. Witnesses are

only allowed to give statements about their own perception and experience. Their statements cannot be based on opinions or assumptions. In practice, the fact that written documents can be used means statements are rarely given at trial, but that the case file is often relied upon. In the case file the statements of witnesses and suspects are written documents. Furthermore, there are also certain minimum requirements that have to be met for the evidence to be sufficient to convict the suspect. For instance, the statement of one witness is insufficient to prove the suspect's guilt. Similarly, a confession has to be corroborated (Verbaan, 2016). Despite these rules about what evidence is to be used, the question remains how the evidence is used to ultimately reach a decision on whether or not the suspect committed the crime he is accused of.

2.1.1. Reasoning with Evidence

As several theories on legal decision-making have been extensively discussed in Chapter 1, they will only be briefly described here. For instance, Pennington and Hastie (1992) suggested that the extent to which the evidence forms a causal and plausible story can affect the decision-making process. The theory of anchored narratives (Wagenaar et al., 1993) proposes that the individual evidence in a case can be seen as substories. These substories then need to be anchored. The judgment as to when the evidence is sufficiently broken down and can be anchored is left to the discretion of the judge (Wagenaar et al., 1993). Simon (2004) applies coherence-based reasoning to legal decision-making. According to that theory, cognitively complex tasks are reconstructed into easier ones. Coherence is achieved when the alternative that has been chosen by the decision-maker is perceived as having strong support, while the rejected alternative has weak support. When both alternatives have equal support, the model is thought to be incoherent. Despite support for some of these theories on the reasoning process of judicial decision-makers, the process remains elusive. As most of the theories are prescriptive, they are inherently hard to test. It is not always clear from the final decision on guilt how evidence was considered, and judges themselves may

be subject to influences they are unaware of, thereby limiting their own insight into the reasoning process (Irwin & Real, 2010).

Despite not being able to determine exactly how decisions are made, there are some aspects of reasoning that have been argued to be of vital importance for determining whether a hypothesis is true. According to Popper (1957), in order to reach a conclusion about a hypothesis, one needs to try to both confirm and falsify the hypothesis. Confirmation alone is insufficient, as one instance contrary to the hypothesis would be able to disprove it. However, if several serious attempts at disproving the hypothesis have been unsuccessful, it becomes more likely that the hypothesis is true. The classical example given here is the hypothesis that all adult swans are white. If you are claiming that all adult swans are white and supporting that hypothesis with white swans that have been seen, you are engaging only in confirmation. The hypothesis is supported, but could easily be disproven by a single sighting of a different coloured swan. Therefore, rather than looking only for white swans to confirm the hypothesis, one should also try to find different coloured swans to disconfirm the hypothesis. If unsuccessful in finding swans of a different colour, it becomes more likely that the hypothesis that all adult swans are white is true.

The same rule of reasoning can also be applied in legal decision-making. In trying to determine whether the suspect committed the crime he is accused of, the investigation, and ultimately the court, should not only try to find evidence that confirms that the suspect committed the crime. Attempts should also be made to disprove that the suspect committed the crime. When these attempts fail, it becomes more likely that the suspect is indeed guilty. Such attempts can be made for instance by checking a suspect's alibi, or through the consideration of alternative scenarios. If there are alternative scenarios of who committed the crime, a thorough investigation into the guilt of one of the alternative suspects can be considered an attempt to falsify the scenario for the main suspect. Furthermore, evidence that does not support the guilt of the suspect should also be considered. For instance, if a witness had a clear view of the perpetrator,

but does not identify the suspect, that can be considered contradictory to the hypothesis that the suspect is guilty. Of course, other possible reasons why the witness did not identify the suspect should be excluded, such as a change in the suspect's appearance or an extreme delay between the crime and the identification procedure. Nevertheless, a non-identification of the suspect should not simply be disregarded (Sauerland et al., 2016). If several witnesses saw the incident under similar circumstances, and one witness identifies the suspect as the perpetrator but the other does not, then it becomes more likely that the suspect is not the perpetrator than if the viewing conditions differed for the witnesses, or the identification procedures were conducted differently. These should then be analysed to determine which outcome of the procedure has more evidential value (Sauerland et al., 2016). Simply disregarding a non-identification would indicate an emphasis on confirmation, as a non-identification could support an alternative scenario to the guilty scenario. When considering evidence, one should not only look for evidence that supports one of the scenarios, but should rather try to find evidence that can discriminate between the different scenarios.

However, there is one aspect of human reasoning that might make the process of trying to falsify more difficult than it may initially seem. This is due to what is known as confirmation bias: the tendency to search and interpret information in line with an existing belief, while paying disproportionately less attention to information that might contradict that belief. Confirmation bias is a well-documented phenomenon and has been demonstrated in various experimental settings (e.g., O'Brien, 2009; Kassin et al., 2003; Dror et al., 2005). Researchers have also looked at confirmation bias in the context of reasoning with evidence (e.g., Rassin, 2018; Ask & Granhag, 2005). According to Findley and Scott (2006), the criminal justice system is particularly affected by the excessive focus on evidence against the suspect and the disproportionately little weight given to evidence for the suspect's innocence. They describe a number of cases to support their argument. Due to the difficulty in identifying the reasoning process that led to a decision, determining whether confirmation bias affected a case remains

challenging. In experimental research, confirmation bias is often measured in terms of participants' selection or rating of the information presented, sometimes combined with their perception of guilt (e.g., O'Brien, 2009; Jonas et al., 2001; Ask & Granhag, 2005). When translating that method to a judicial decision, it would require an analysis of the evidence used, and not used, by the court to reach its final decision on whether or not to convict the suspect.

2.1.2. The Current Study

The best way that is practically possible, by which insight into the court's reasoning can be gained, is through the decision supported by reasons that the court has to provide according to the Dutch Code of Criminal Procedure. In the Netherlands, if the court chooses to convict the suspect, it is required to provide facts and circumstances that support that decision, as well as to include the evidence for the decision. If the decision deviates from explicit, motivated points raised by the prosecution or defence, the court also has to explain that deviation (art. 359 CCP). Nevertheless, it has become clear from rulings by the Supreme Court that the motivation need not be an accurate reflection of the evidence considered by the court, but rather, it has to include the information that the decision can rest upon (Reijntjes & Reijntjes-Wendenburg, 2018). Therefore, the insight gained into what the court actually considered remains limited, although the written decision is the best representation of the reasoning process and its outcome that is available.

In this case study, the aim was to explore the reasoning process explained by the court in the reasoned decision, using the evidence in the case file to determine whether attempts at both confirmation and falsification were made. Furthermore, the evidence used by the court to find the defendant guilty was examined for its ability to discriminate between different scenarios. For instance, evidence that could fit with all the proposed scenarios is not able to discriminate between the different scenarios, and should therefore not be used to determine guilt or innocence (Wagenaar et al., 1993).

The case at hand was part of the Project Reasonable Doubt (Amsterdam Lab for Legal Psychology, n.d.), meaning that the individual convicted in the case maintains he is innocent, but his conviction has exhausted all legal remedies possible in The Netherlands. In this case, a young man was stabbed to death during carnival celebrations. The choice for this particular case was made because the Advocate General of the Supreme Court noted that the conviction rested on a 'thin evidence construction'. Nevertheless, when reading the decision, it becomes clear that several scenarios were considered by the court. It is therefore believed that the case serves as an interesting example of reasoning with evidence, where a decision supported by both the court of first instance and the court of appeal can still be considered 'thin' by the Supreme Court. Before exploring the reasoning in the case, a brief description of the case is given. Subsequently, the verification and falsification attempts by the court that have been identified in the have been critically examined. By doing so, the aim was to explore whether falsification was used, as well as how it affected the final decision to convict the suspect. As the court of appeal for the most part agreed with the court of first instance, the current analysis focused on the decision made by the court of first instance.

2.2. The Case

In the case that will be used for this case study, a young man was convicted for murdering another young man. He was found guilty of murder by stabbing. Firstly, a neutral description of the case will be given. As will become evident, it is unclear what exactly happened as different witnesses provide a different version of the events. However, for the events preceding the stabbing, there appears to be consensus on what, broadly speaking, happened, although the details remain disputed by the various people involved. The version of events described below is based on the court's reasoned decision (Utrecht Court, 2012).

On the 18th of February 2012, 26-year old Thijs was stabbed during carnival celebrations in a small town in the Netherlands. He had been celebrating carnival

with his friends earlier that evening. The defendant, Remco, 23, had also been celebrating. Earlier that evening, Sebastiaan, a friend of Thijs had been in a fight with Kevin, a friend of Remco. Thijs had seen this fight and had gotten involved, and had ended up fighting Kevin. Kevin then ran away, at which point Thijs, together with his brother, chased Kevin. Remco saw that, and when the chase came to a halt by a parked car, Remco shouted at Kevin that he should run. At that point, Thijs and his brother became focused on Remco. Kevin managed to run away and met up with his friend Boy. They initially wanted to go back to find Thijs and his brother, because Kevin was still angry, but they could not find Thijs or his brother.

After Kevin ran away, Thijs tried to get back in contact with Kevin by grabbing Remco and asking him to call Kevin. Thijs and his brother hit Remco and stole his mobile phone. They told Remco that he would get his phone back if he could get Kevin to come back. Remco then ran to his house, which was not far from where the fight happened. At this point, nobody followed him. When he got home, he realised he wanted to get his phone back, so he returned to where the fight had taken place. There, he found Thijs and his friends, and shouted at them that he wanted his phone back. When Thijs and two other guys turned towards Remco, he started running away immediately. He was chased by Thijs, Sebastiaan, Thijs's brother, and the girlfriends of Thijs and Sebastiaan. The group was running in that order, according to camera footage of the event.

Less than two minutes after Thijs was last seen on the CCTV recordings, the first call was made to the emergency services about the stabbing. Officers responded to the call, and found Thijs, half lying against the wall in an alley, two doors down from where Remco lived. He appeared to have been stabbed. Thijs was taken to hospital where he died later that evening.

When the investigation began, Kevin was initially suspected of killing Thijs, as he had been in a fight with him earlier that evening. Kevin then accused Remco, who was arrested the following day. According to Kevin and Boy, Remco had later that evening confessed to them that he had stabbed Thijs.

When interviewed by the police Remco consistently denied that he stabbed Thijs. Following the police investigation, the Court of First Instance considered it proven that Remco had murdered Thijs. In its decision, it mentioned several things it considered incriminating for Remco, as well as several scenarios of other perpetrators that it did not find likely. The court also specified that it considered possible influences on the witness statements, for instance from the media or from rumours, and that drugs and alcohol could have affected the memories of witnesses. Therefore, the court stated that only statements that were supported by other statements or objective facts were used to determine what was proven.

2.2.1. Attempts at Confirmation

The court included several pieces of information that it considered to be incriminating for Remco, and thus, confirmatory for the scenario in which he is the perpetrator. The court sums up the evidence that supported the decision which included: the fact that he was present at the crime scene, the fact that he had been in a fight earlier that evening and therefore had a motive, and the fact that he made a confession containing knowledge that was unique to the perpetrator to three friends that evening.

In critically examining these pieces of evidence, the questions raised include whether the evidence used by the court is supported by other information in the case file, and whether it can discriminate between the scenarios in which Remco killed Thijs or the scenarios in which someone else killed Thijs. Firstly, the fact that Remco was at the crime scene. Although several witness statements, as well as Remco's own statement, indeed confirm that he was at the crime scene, there is no evidence that Remco was at the crime scene at the exact moment that Thijs was stabbed.[1] According to Remco himself, he did not stab Thijs and did not see who did. He continued running into the alley until Sebastiaan caught

1 According to the pathologist, it is unlikely that Thijs was stabbed before the chase. None of the witnesses mentioned Thijs being stabbed prior to the chase. One witness described him as wearing an outfit with red, although his costume was entirely white.

up with him, at which point the two men started fighting. Remco's statement is also supported by Sebastiaan's statement, who also states that he ran into the alley and did not see Thijs. Sebastiaan was a good friend of Thijs, and therefore it appears unlikely he would lie in order to protect Remco.

The order in which the men were running is also significant in relation to the evidence that Remco was at the crime scene. According to Sebastiaan, he was running directly behind Remco, and Thijs was running behind him. That places Sebastiaan between Thijs and Remco, making it impossible for Remco to have stabbed Thijs without Sebastiaan seeing it. Sebastiaan's statement therefore undermines the incrimination provided by the fact that Remco was at the crime scene. Furthermore, considering that no one saw Remco stab Thijs, his mere presence at the crime scene does not discriminate between the scenario in which he is the perpetrator and the scenario in which someone else is the perpetrator, especially considering the fact that multiple people were present at the crime scene in a very short period of time.

Secondly, the court mentioned the fact that Remco was involved in a fight with Thijs earlier that evening, which it considers a motive for the murder. Although used as confirmation of Remco's guilt, the fact that Remco had a fight with Thijs earlier in the evening does not discriminate between a scenario in which he did kill Thijs and a scenario in which he did not kill Thijs. Besides the fact that the stabbing did not occur during or directly after the fight, Thijs had also been in a fight with Kevin that night. Moreover, Remco did not seem to know Thijs. When first questioned by the police, Remco initially seemed to think they were asking him about Sebastiaan, a friend of Thijs' with whom Remco had fought after the chase. Remco was also not involved when the fight initially started, as he only became involved because he told Kevin to run away. Although it could be argued that Remco was angry at Thijs for taking his phone, it could also be argued that Remco had no interest in an altercation with Thijs, seeing as he was running away from Thijs and his friends.

Perhaps the most important evidence used by the court to confirm that Remco was guilty were the witness statements made by Kevin and Boy. It is here important to note that the court used only the verbatim report of the first statement by Kevin.[2] Its reasoning for that was that he had not yet had access to the case file, and he had been arrested only three hours after the stabbing. Therefore, the information provided by Kevin during that interview could not have come from other sources. The court also notes that at the time that Boy made a statement, Kevin was still detained in isolation, which means he could not have spoken to Boy. Kevin and Boy claimed that Remco had confessed to stabbing Thijs the night of the attack. Based on the statements made by Kevin and Boy, the court concluded that Remco had said that he had lost "a/ the knife". Another friend, Barry, was also present during the conversation described by Kevin and Boy. Barry initially denied that Remco had said anything about a knife but when questioned by the investigative judge, Barry said that he heard either Remco or Kevin, but he did not know who, say that he had stabbed Thijs. The court considered this confirmation of the statements made by Remco and Kevin. In his statement, Boy also said that Remco had stabbed Thijs twice. Boy made that statement before the coroner's report, confirming Thijs was stabbed twice, had been added to the case file. The court therefore considered it guilty knowledge, which he must have gotten from Remco.

There are a number of issues that arise when taking a closer look at Kevin and Remco's statements about the discussion. The court states that the conversation between the four men took place in Remco's bathroom. However, in his first statement, which is the one used by the court, Kevin said the conversation took place in the garden. That contradiction with the others' and his own later statements undermines the validity of his statement, as only one of these versions can be true. Nevertheless, the court saw no problem in attaching extensive weight to Kevin's statement, in part due to the fact that they had watched the

2 We do not have the verbatim report of that interview. We only have a two-page summary of what was said by Kevin.

recordings of the interview, and had concluded that Kevin's statements were voluntary and authentic.[3] Moreover, the suggestion by the court that Kevin and Boy's statements were independent can also be questioned. Kevin and Boy were very good friends. Although they did not have any contact after Kevin was arrested, they both state that they left Remco's house together that evening. It is therefore possible that they discussed either the stabbing or the conversation they had just had with Remco in the time they were together.

Another issue is the contradictory statements initially made by Barry. In his interviews, Barry gave a very different version of the conversation between the four men. According to Barry, Remco said he did not stab Thijs. Both Remco and Kevin were extremely emotional. Remco knew that someone had been stabbed but did not know by whom. Based on Barry's statements, it is also not clear if Remco knew who had been stabbed. Boy had also described a stabbing motion that Remco made, which was again contradicted by Barry, who claimed the light in the bathroom had been off, and so it would have been difficult to see this. Although the light being off might offer an explanation as to why Kevin could not describe the bathroom, it is also another inconsistency between the witnesses. The interviewers threatened Barry with perjury after he continued to deny that Remco had confessed to stabbing Thijs. The statement made by Barry at the investigative judge is also very clear in that he did not know whether Kevin or Remco had said they had stabbed someone, and that he was not sure exactly how this was said. It might also be relevant to note that Barry claims he has a hearing problem, and so he might not have heard everything that was said. Nevertheless, Barry's statement that Remco was sure he did not stab Thijs is a stark contradiction to the statements made by Boy and Kevin, who were more certain about Remco's confession and even gave details of the stabbing. Interpreting Barry's statement as incriminating for Remco seems more difficult, as he was not sure who had confessed. It is therefore difficult to use Barry's

3 We did not have the audiovisual recordings of the interview

statement to discriminate between the scenario in which Remco killed Thijs and the scenarios in which he did not.

The guilty knowledge that Boy seemed to have, namely that Thijs was stabbed twice, is also seen by the court as incriminating for Remco. The court excluded the possibility that the perpetrator knowledge came from Kevin, as they had excluded Kevin as a suspect. However, the reasoning that someone is not a suspect and therefore could not have had perpetrator knowledge seems tautological. Seeing as both Remco and Kevin were present in the bathroom, the perpetrator knowledge could have come from either of them. Therefore, the fact that Boy heard the perpetrator knowledge during that conversation does not discriminate between the scenario in which Kevin is the perpetrator and the scenario in which Remco is the perpetrator. However, seeing as the court already considered the former scenario to have been excluded, they considered the fact that Boy possessed guilty knowledge as confirmation of Remco's guilt.

One last factor that should be considered when determining whether Boy and Kevin's statements really did confirm Remco's guilt is the fact that Kevin was the initial suspect in the case, and he pointed the finger at Remco. He therefore had an incentive to blame Remco. Boy was a good friend of Kevin's, and Barry was a good friend of Remco's. It is possible that the men lied to protect themselves and their friends. Whereas the court assumed that was the case for Barry, it is not evident from their reasoned decision that they considered the same argument for Boy's statements. It therefore appears there was a tendency to try and confirm Remco's guilt, whereas there was no attempt to try and disprove the statements made by Kevin and Boy. After having looked at several factors to determine if these statements could be disproven, it appears their support for the scenario in which Remco is guilty is not as clear as suggested by the court's written decision.

One could argue that despite the evidence individually not confirming Remco's guilt, the evidence overall can still be convincing. Although that may be true, it should still be considered whether the evidence overall is able to discriminate between the alternative scenarios of what happened. The court

also used the evidence in the case to consider several scenarios in which someone other than Remco was the perpetrator. That can be considered a form of falsification, seeing as support for the scenario in which someone else killed Thijs would contradict the scenario in which Remco killed Thijs.

2.2.2. Attempts at Falsification

The main attempt at falsification by the court can be found in the alternative scenarios that it considered. The court outlined four different scenarios: a) the friends and family who were running after the victim stabbed him, b) a person who remained unknown and was chasing Thijs stabbed him, c) Kevin stabbed Thijs, and d) Remco stabbed Thijs.

The first scenario could easily be excluded by the court. The case file contained no indication for that scenario, and the court considered it absolutely unlikely. It therefore did not need to be discussed further. Although the court's evaluation of that scenario is easily understood, it is unclear why this scenario was raised at all by the court. As will be discussed later, there were other, more likely, scenarios that the court would have benefitted from excluding.

The second scenario was also considered unlikely by the court. Due to the stab wounds, which were both at the front of the body and upward, the court considered it unlikely that an unknown individual had stabbed Thijs during the chase. According to the court, the injuries suggest there was a face-to-face confrontation during which Thijs was stabbed. If such a confrontation had taken place, it must have happened around the start of the chase. No one mentioned such an encounter in their statement. The court also stated that a possible motive for an unknown individual was missing. It should be noted that there also seemed to be no focus on finding such a motive during the police investigation. The court therefore decided that that scenario, which found no support in the case file, did not need to be considered further.

Although we agree that the scenario in which an unknown individual stabbed Thijs is unlikely, it could be argued that the consideration of the lack

of motive for an unknown individual is a strange consideration. In order for a motive to be clear, the individual would probably need to be identified. Seeing as the motive is inherently linked to a possible perpetrator, using the lack of a motive as reason to exclude an unknown individual is not a strong argument. Furthermore, as discussed earlier, what the court considered to be a motive for Remco, namely his earlier fight with Thijs, also did not discriminate between the scenario in which Remco killed Thijs and the scenario in which he did not.

The third scenario considered by the court is the scenario in which Kevin stabbed Thijs. The court excluded that scenario. The court acknowledged that Kevin was angry and full of emotion. However, as Kevin had run away from Thijs earlier that evening, the court seemed to think it unlikely that he was involved in Thijs's death. Kevin tried to find Thijs again later in the evening, but says he could not find him. What the court considers to be exculpatory for Kevin is similar to the events as described by Remco. Remco had also run away from Thijs earlier in the evening. He returned to where Thijs and his brothers were to try and retrieve his phone. When the group turned towards him, he again ran away. The court did not consider Remco running away to be exculpatory in the same way as they did for Kevin.

The court also used the fact that Kevin was with Boy and the fact that no one had seen him as grounds to exclude the scenario that Kevin stabbed Thijs. According to the court, Kevin and Boy were at the windmill located near the city center around 7.45/7.50 PM. The court therefore seemed to consider that Kevin had an alibi for the time at which Thijs was stabbed. However, based on CCTV footage and the first call to the emergency services, we know that Thijs was stabbed between 7.41 PM and 7.43 PM. It would only take a minute to run from the windmill to the crime scene. Furthermore, Kevin's location was determined based on the telephone records and the witness statements. Considering the close location of the crime scene to the windmill, the phone records are not informative about where exactly Kevin was located (van Bree,

2015). It therefore seems problematic to use Kevin's whereabouts as an alibi to exclude him as a suspect.

Lastly, the court also considered that none of the witnesses saw Kevin at the crime scene. It is difficult to determine how important it is that Kevin was not seen. For instance, as we do not know the exact amount of time between Thijs's arrival in the alley and his brother's arrival in the alley, we can only speculate about what his brother should have seen had either Remco or Kevin stabbed Thijs. It is also unclear what Sebastiaan should have seen, as we do not know exactly where he was running in the chase. As described earlier, he states he was running in between Thijs and Remco. Although that makes it difficult for Remco to have stabbed Thijs, it makes it more likely that Sebastiaan would not have seen Thijs getting stabbed behind him. It is also possible that Sebastiaan was running behind Thijs. In that case, the fact that Sebastiaan did not see Thijs, if he was stabbed by Remco, remains strange. One possibility is that Sebastiaan was focused on Remco and therefore did not see Thijs who had been stabbed. That could be explained by inattentional blindness, which causes individuals who are focused on one thing to not perceive or remember other things around them (Simons & Chabris, 1999). That could also mean that Sebastiaan may not have noticed other people if he were exclusively focused on Remco. Kevin had lost most of his costume during the earlier struggle, and so he would not necessarily have been easily noticeable by Sebastiaan. Of course, none of the considerations described here prove that Kevin was at the crime scene. However, it seems as if the court may have too easily accepted the support for Kevin's innocence.

Although the reasons given by the court to exclude the suspect might not be convincing, the court's explanation of the alternative scenario nevertheless indicates an openness to the possibility of alternative perpetrators. However, it appears that, unlike for the scenario in which Remco was the perpetrator, they did not try to confirm the scenario in which Kevin was the perpetrator. Although it is possible that attempts to confirm the scenario were simply not mentioned in the court's decision, the case file also did not offer much of an explanation as to

why some of the incriminating information did not lead to further investigation or consideration. For instance, Kevin was also present at the conversation where guilty knowledge was exchanged. Barry was not sure who said they had lost the knife during the conversation in the bathroom. That appears to have been disregarded by the court, who interpreted Barry's statement as supportive of the accusations towards Remco.

During the search of Kevin's house, the police also found a drawing of the crime scene. Kevin's mother told the police that Kevin had made the drawing as a clarification for his father. According to Kevin, he met up with Remco and the others in Remco's garden and he did not go to the crime scene. Furthermore, no one mentions discussing where exactly Thijs was stabbed. The drawing therefore raises some questions. However, the police only asked Kevin's mother about the drawing. The explanation that Kevin had made the drawing as a clarification for his father does not explain how Kevin knew exactly where Thijs had been found. Although the drawing could have offered support to the scenario in which Kevin stabbed Thijs, there was no mention of it in the consideration of the scenario that was outlined in the court's decision. Although we again cannot conclude that it was therefore also not considered, the case file also lacks any further investigation of why the drawing was made, including asking Kevin himself.

In addition, several witnesses also stated that Kevin was known to carry a knife. Although that of course does not prove his involvement in the death of Thijs, the police seemed to pay no attention to it at all. In contrast, when they discovered a 2-year-old chat conversation about a knife from Remco, they investigated what had happened to the knife since then. Although neither Kevin's carrying of a knife, nor Remco's conversation about a knife connect them to the murder of Remco, the fact that Kevin's history did not seem to be considered by the court could indicate a lack of interest for confirmation of the scenario that Kevin stabbed Remco. The lack of investigation into Kevin as a suspect was also not considered by the court, which supports a lack of interest in falsification in their reasoning process.

Although there is no strong evidence against Kevin, similarly to the weak support for the scenario that Remco killed Thijs, there are a few things that support the scenario that Kevin is guilty of killing Thijs. A clear difference in how the scenarios were investigated by the police and considered by the court can be seen. Whereas there seemed to be a preference for confirming the scenario in which Remco is guilty, disconfirming the scenario in which Kevin was the perpetrator seemed to have the preference of both the courts and the police.

In addition to the consideration of Kevin's scenario appearing to be influenced by a belief in Remco's guilt, there was also another scenario which the court briefly mentioned, but did not consider likely. That is the scenario in which Sabrina, Remco's girlfriend, killed Thijs. According to the court, she could be excluded as a suspect because she was not present during the conversation where the perpetrator spoke about the murder, and because she did not have a motive. Based on the statements by the people present at that conversation, it is unclear what exactly was discussed and by whom. The assumption that Sabrina could not be the perpetrator because she was not present at the conversation where the perpetrator talked about stabbing Thijs, implies that the perpetrator is known to be one of the people present at the conversation. An alternative explanation would be that one of the people present at the conversation could also have heard perpetrator knowledge from somewhere else.

The fact that Sabrina did not have a motive also seems insufficient to exclude her as a suspect. From the case file, it seems Sabrina knew about the altercation. She saw Remco running towards the house, and ran towards him to help. She was in the house at the time, and had the opportunity to grab a knife. More importantly, Sabrina also had the opportunity to stab Thijs. Whereas the court used Remco's presence at the crime scene as support for his guilt, they did not consider the same for Sabrina. Sabrina herself says she pulled Thijs back during the chase. She therefore had the opportunity to stab him, or should have seen who did. According to the timeline, she was the last person to see Thijs before he was found by his brother. Thijs's brother and girlfriend did not see Sabrina

when they turned into the alley where Thijs was found. Due to the lack of information about the exact distance between the people in the chase, we cannot conclude how much time passed between Sabrina pulling Thijs away and the others arriving. Seeing as no one else describes Sabrina pulling away Thijs, we have to rely on her own statement.

Overall, the consideration of the alternative scenarios offered by the court seemed to disregard some of the evidence in the case file. Although this does not necessarily mean they also did not consider that evidence, it results in their consideration appearing to be biased by their decision that Remco was guilty. Moreover, the case file also seemed to offer insufficient information to unquestionably exclude some of the scenarios. That finding is also in line with the conclusion by the Advocate General, which was provided as part of the confidential case file, and states the evidence construction for the conviction was 'thin', especially when considering the alternative scenarios that arise from the case file. Therefore, further investigation into these scenarios, or the demand for that investigation by the courts, would likely have been beneficial and would have been better proof of falsification attempts.

2.3. Discussion

When determining whether the court attempted falsification on the basis of the written decision, it must be remembered that the written decision is probably not a good representation of what the court considered. The court only has to explain why it considers something to be proven. Nevertheless, the court in the current case explained how alternative scenarios were excluded. Based on the written decision, as well as the case file, the reasoning of the court seems to have been biased towards Remco's guilt, which resulted in questionable interpretations of the evidence. Of course, the function of the reasoned decision is to explain the decision made. So, considering the conviction of Remco, it is not surprising that the reasoned decision would focus on Remco's guilt and why other perpetrators are not likely. However, when taking into account the

evidence in the case file, the reasoned decision seems to offer an insufficient explanation as to why the court was convinced of Remco's guilt.

One difficulty the current study encountered was in determining what evidence or scenarios were actually considered by the court. The written decision is known not to be a valid representation of the decision-making process. It suffices for the written decision to be an indication of what the decision could rest on (Reijntjes & Reijntjes-Wendenburg, 2018). It can be argued that in order for the court's decision to be understood on the basis of the written decision, it would be beneficial for it to be a reflection of the reasoning process rather than an explanation of the decision. It would then be possible to know not only what the court considered important, but also what they did not consider important, and why. The court may have attempted falsification in their reasoning, but if this failed, it would likely not be visible in the reasoned decision, as it would be a necessary element to support Remco's guilt.

When trying to understand the decision made by the court, we ideally want to reconstruct the decision-making process. The distinction between the actual reasoning process and the representation of the process in the written decision can also be related to the distinction between context of discovery and context of justification. Context of discovery refers to the process of constructing a theory, whereas the context of justification refers to communicating, or justifying, the theory to others (Swedberg, 2011). Several interpretations of the distinction exist, for instance between the process of discovery and the methods of justification (Hoyningen-Huene, 2006). Some also argued that they are temporally separate processes (Hoyningen-Huene, 2006), which could also apply to the decision made by the court, where a decision is made before the writing down of the decision. Studying the emergence of new theories, or the context of discovery, is difficult in practice (Swedberg, 2011). Therefore, although it would in any case be difficult to determine how a theory about the guilt of the suspect emerged, the written decision can only offer insight into the context of justification, and the manner

in which the court decides what possible versions of events to consider, remains elusive.

When examining what was considered proven using the structure proposed by the theory of anchored narratives (Wagenaar et al., 1993), the conviction appears to be based on an unstable construction. The various substories which were anchored, namely the evidence used by the court of first instance, contain a number of problems. For instance, the statements given by Kevin and Boy are an example of one of the rules that are commonly accepted given by Wagenaar et al. (1993), namely that two witnesses who make the same statement can hardly be wrong. As discussed above, the validity of the statements given by Kevin and Boy can be questioned for several reasons. Nevertheless, the court used the statements to reach the conclusion that Remco confessed to Kevin and Boy, and thus was guilty.

The alternative scenarios excluded by the court might indicate falsification attempts. Logically, the exclusion of alternative scenarios was reliant on disproving the scenarios. However, when considering the evidence in the case file, some of the information used to disconfirm the scenarios seems insufficiently strong to consider the scenario in which Remco was guilty as most likely. It appears as if the court tried mainly to disconfirm the alternative scenarios while trying to confirm the scenario in which Remco was guilty, whereas the value of falsification is in trying to disprove the preferred scenario (Van Koppen & Mackor, 2019). In the current case, that does not seem to have happened: there are several signs of confirmation bias in both the case file and the written decision.

One thing that is illustrated by the current case study is the difficulty in determining whether something can be considered confirmation or falsification. These concepts are difficult to apply in certain contexts. For instance, the consideration of alternative scenarios can in theory be considered an attempt at falsification, but when the focus is on disconfirming these alternative scenarios, it seems to again become an attempt at confirmation of the original scenario. Therefore, it may not always be clear what one needs to do in order to use both

confirmation and falsification properly to decide on whether or not something is true.

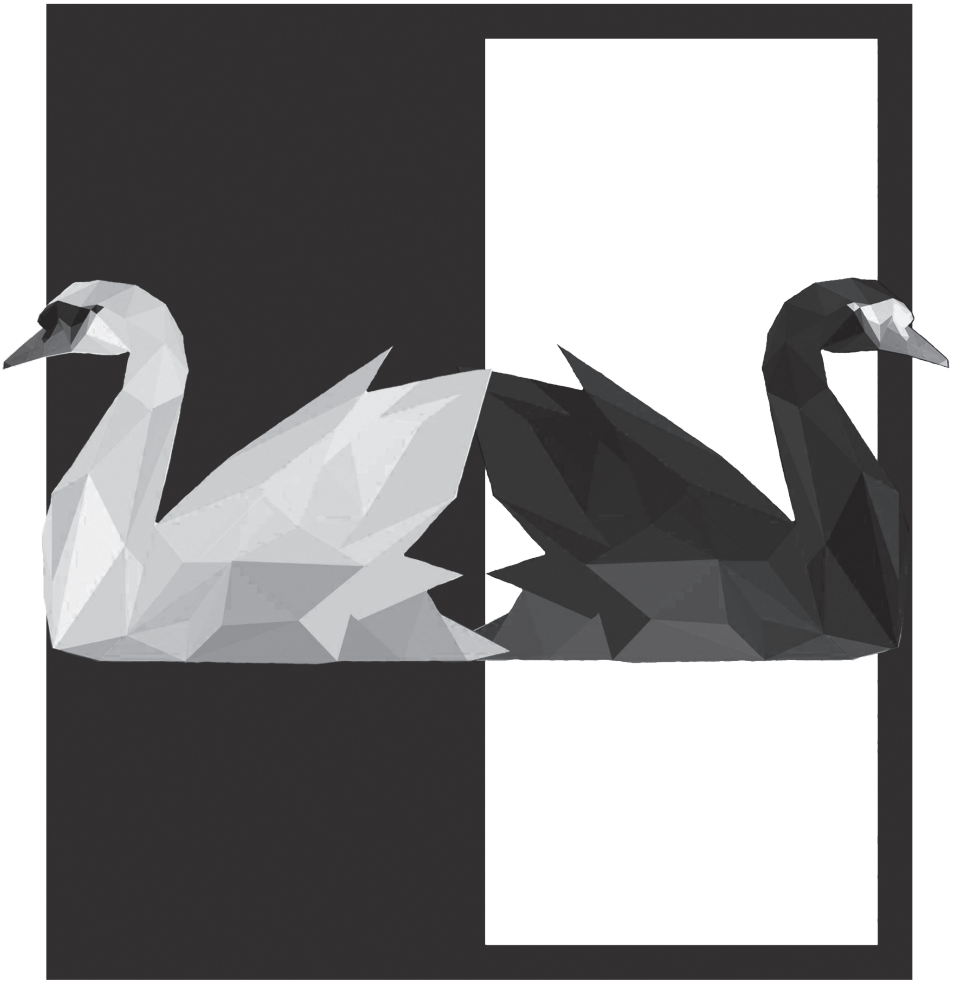
In this case, a discrepancy between the theory and the application in practice also became clear in relation to the use of alternative scenarios, at least based on the analysis of the written decision. The scenario approach relies on the inference to the best explanation. That process uses alternative scenarios which explain the evidence, after which the best scenario is chosen based on comparison of the scenarios (Bex, 2015). The best scenario is considered to be the scenario that offers the best explanation of the evidence out of the proposed scenarios (Van Koppen & Mackor, 2019). Determining which scenario best explains the evidence is a different process than determining which scenario has the most support: one scenario may have plenty of support, but fail to explain the rest of the evidence. Similarly, some evidence may be explained by several scenarios and have no diagnostic evidence, which would not be recognized when reasoning in terms of support for the different scenarios.

Based on the current study, as well as the findings from the case study, it appears that the consideration of alternative scenarios by the court does not necessarily involve the construction of scenarios based on the evidence, nor does it necessarily involve looking for the scenario that the evidence best fits with. Rather, the consideration of alternative scenarios seems to happen in such a way that confirmation (or disconfirmation) is sought for those scenarios that the court feels warrant consideration. The guilty scenario, or charge, can be constructed on the basis of evidence from the case file, that the prosecution is free to choose (Crombag & Wagenaar, 2000). The construction of alternative scenarios is undoubtedly in part also based on the evidence in the case file, although scenarios also seem to be constructed because ruling them out is considered necessary by the court (e.g., the scenario that family members stabbed Thijs). That may also apply when a scenario is raised by the defence. An alternative scenario can only be weighed against the competing scenarios if it has been thoroughly investigated (Crombag & Wagenaar, 2000).

As confirmation bias results in favouring evidence for a preferred scenario, as well as under-appreciation of evidence that may contradict that scenario, simply looking for evidence for or against the scenario is likely to also be affected by confirmation bias, and thus unlikely to protect against bias. In this case study, that also appears to have affected the consideration of scenarios. In trying to find support for the different scenarios rather than looking at how the evidence fits with each of the scenarios, preference seems to have been given to supporting the scenario of Remco's guilt. Therefore, in our analysis, we looked at how the evidence fits with the different scenarios, thereby considering how the evidence could discriminate between the scenarios.

The current study was based on the case file and what the court wrote about the information within the case file. In order to analyse the decision process for the purpose, we made use of the evidence in the case file that could, to some extent, discriminate between the scenarios offered by the court, as well as alternative scenarios that could be derived from the case file. That does not exclude the possibility that there may be other scenarios which others would consider more likely. As became clear, our interpretation of the evidence, and its subsequent diagnosticity, was not always in line with the court's interpretation. Although the arguments given in this case study were considered to be the most important elements for the purpose of the study, it remains possible that others may have further additions to our analysis, or may have different emphases than in our analysis.

The findings of the current case study indicate that the court tried to use falsification by considering alternative scenarios, but may not have done so properly. As it cannot be excluded that falsification attempts were also simply not included in the written decision, the conclusions we can draw are limited. It is important to note that an incomplete representation of the thought-process could result in the decision seeming unreasonable. An improved documentation of the considerations made in the decision-making process could counter such an impression.



CHAPTER 3

JUDGES' VIEWS, EXPERIENCE AND KNOWLEDGE OF FALSIFICATION: AN EXPLORATORY SURVEY

Abstract

The role of falsification in legal decision-making remains largely unknown. Falsification is often disregarded due to the well-established phenomenon of confirmation bias. In order to gain insight into how falsification is regarded and applied by judges in practice, we distributed a questionnaire among criminal law judges in The Netherlands. The questionnaire consisted of three parts: knowledge, application, and experience. In the first section, participants were asked whether they agreed or disagreed with statements about falsification and the consideration of alternative scenarios. In the second section, participants were asked to complete a number of tasks, such as a variation of the Wason selection task. In the final section, participants were presented with open-answer questions about their experience in practice. Due to a low response rate, the answers of the questionnaire were supplemented with semi-structured interviews with criminal law judges. Based on the collected responses, judges seem to understand the importance of falsification, although the extent to which they apply it, or consider that their task, varies considerably.

3.1. Introduction

In several European legal systems, the court has to determine whether or not the suspect is guilty of the offence he is charged with. In The Netherlands, the judge can convict the suspect if he is convinced based on sufficient legal evidence that the suspect is guilty. In case he is not convinced, or there is insufficient evidence, the judge must acquit the suspect. The judge is free to decide what evidence to use (within the types of evidence regulated by the Code of Criminal Procedure) and what value to give to the evidence (Stevens, 2014). Little is known about how judges select and value evidence, although there appear to be large differences between judges (Buruma, 2009). In the current study, a survey and interviews were conducted with criminal law judges in order to gain insight into their use of falsification and consideration of evidence, particularly evidence that contradicts the charge against the defendant.

It has been argued that, in criminal law proceedings, there is an inherent risk of confirmation bias affecting decisions (Findley & Scott, 2006; Rassin et al., 2010). Confirmation bias is a human tendency to search for, and interpret, information in line with an existing belief, while paying disproportionately less attention to information which contradicts that belief (Wason, 1968; Nickerson, 1988). That risk arises in part due to the composition of the case file. In The Netherlands, the case file is likely to contain mainly incriminating information (Crombag, 2017). In preparation for the trial, the judge will read the case file or a summary of the case file. It is then likely that the judge's initial belief will be based on the case file. Such an initial belief can start a cognitive process resulting in confirmation bias. Namely, once that belief has been formed, being presented with contradicting information, for example during the court session, may cause cognitive dissonance, a mental state associated with an uneasy feeling (Festinger, 1957). In order to reduce dissonance, belief perseverance can arise, whereby the initial belief is adhered to by rejecting contradicting information. Confirmation bias then further helps to maintain the initial belief by emphasising, and interpreting evidence, in such a way that the initial belief can be confirmed

(Nickerson, 1998; Kassin et al., 2013). Thus, confirmation bias could negatively affect judges' reasoning by causing an excessive focus on evidence that supports their initial belief. As described above, that belief is likely to be that the defendant is guilty. Similarly, confirmation bias can also cause judges to be less receptive to alternative scenarios (Rassin, 2010). If the investigation has focused on a wrong suspect, confirmation bias in judges' reasoning can contribute to a miscarriage of justice.

A lack of attempted falsification can be considered a manifestation of confirmation bias. Falsification refers to searching for facts that disconfirm the existing belief. Popper (1959) addressed the need for falsification rather than merely confirmation; a finding that confirms a theory can only support the theory until it is overthrown by a finding that disconfirms the theory. The fact that a theory has not yet been falsified cannot lead to the conclusion that it must be true. However, if several serious attempts at falsification have failed, the likelihood that the hypothesis is true increases (Crombag et al., 1992).

In the context of criminal law proceedings, when determining whether or not the suspect is guilty of what he is accused of, it should be determined whether the charges are the most likely scenario of what happened. The charges can for instance be compared to the scenario proposed by the defence by using the available evidence. Whereas some evidence may discriminate between scenarios, other evidence may be in line with several scenarios (Van Koppen, 2011; Van Koppen & Mackor, 2019). Subsequently, a focus on evidence that supports the charges can cause alternative scenarios to be insufficiently considered or evaluated. Based on experimental research, it appears that merely formulating alternative scenarios is not sufficient. The evidence should be carefully considered under each of the scenarios (O'Brien, 2009; Rassin, 2018).

To reduce the risk of confirmation bias affecting legal decision-making, it is important that judges are aware of the need to attempt falsification. In other words, they should not only look for evidence that confirms the charge, but should also look for exonerating evidence, and should consider alternative

scenarios of what happened (Van Koppen, 2011). In a previous study (Van Veldhuizen et al., 2019), 57 judges gave their opinion in response to two statements relating to falsification. The vast majority of respondents (94.9%) agreed that the consideration of alternative scenarios falls under the judge's job. Similarly, 88.1% of the judges agreed that to test an idea about what happened, it is important to try to find evidence that contradicts that idea. Therefore, judges seem to understand that it is important to attempt falsification. In contrast, in many known miscarriages of justice, these attempts were absent or failed, as evidence that supported the suspect's innocence has often been overlooked, for example in the Schiedam Park murder (Rassin et al., 2010). Similarly, according to the case study described in the current thesis (Chapter 2), using falsification in a case may be difficult despite, judges' willingness to do so. There may also be differences between individual judges in their interest and ability to use falsification.

3.1.1. Current Study

Despite the finding that judges seem to understand the need to investigate alternative scenarios, the question remains whether they have an adequate understanding of falsification, and are able to use it in practice. Based on the theory described above, it is clear that falsification is a key element in determining whether or not a theory or hypothesis is true. Although falsification has been suggested to be important in legal decision-making (e.g., in the scenario approach; Van Koppen & Mackor, 2019), to our knowledge, the perspective of judges has not yet been explored. In order to relate the theoretical understanding of bias and falsification to the application by judges, we explored judges' perception of, and experience with, applying falsification in practice using a questionnaire. In the questionnaire, judges were presented with statements about falsification, in the context of legal decision-making, in order to test their understanding and gain insight into what they consider to be important. Respondents were then given three tasks in which falsification should be applied

in order to determine whether, beyond understanding the concept, they could also apply falsification. Lastly, participants answered open questions about their use of falsification in practice. In sum, the questions we aimed to answer were whether judges understand falsification, whether they can apply it, and how they try to apply it in practice.

3.2. Methods

3.2.1. Respondents

The questionnaire was initially sent to 36 judges who expressed interest in taking part in further research in the study by Van Veldhuizen et al. (2019). These judges were all criminal law judges. Furthermore, a few other judges within our extended personal network were also contacted personally to ask if they wanted to take part, and whether they knew any other judges willing to participate. Although we received positive responses from judges willing to distribute the survey, we do not know how many judges were asked to participate.

In the end our response rate was quite low, especially for the final section. For the first section, 36 respondents answered all questions, for the second section, 27 respondents completed the tasks. Nineteen respondents answered the open-questions in section 3. Two of these respondents were removed as they misinterpreted what was meant by falsification. As section 1 and 2 did not include the term 'falsification', it was decided to keep their answers for these sections. Respondents' experience ranged from having been a judge for only a few months to 33 years of experience, with an average of 14.9 years of experience. The average age of respondents was 52.4 years (SD = 10.5). Of all respondents, 48.7% was male and 41% was female. Three of the respondents did not want to report their gender.

Due to the low sample size, we decided to conduct additional semi-structured interviews based on the open questions in the final section. For that, we interviewed three criminal law judges. The aim of the interviews was to have a more in-depth discussion about the concepts addressed in the open questions.

For instance, some of the themes that were identified in the survey responses were expanded on by the judges who were interviewed. Unfortunately, we did not get permission to interview more judges from the Raad voor de Rechtspraak, despite knowing more judges who were interested in participating.

3.2.2. Materials and Procedure

The questionnaire was sent to the judges in an individual email. A link to the questionnaire was provided. The online survey platform Qualtrics was used for data collection. After completing the questionnaire, respondents were thanked for their participation and were asked whether they wanted to be informed about the outcome.

The questionnaire started by thanking respondents for their interest in taking part. The aim of the questionnaire was described as gaining insight into the reasoning process used by judges. After starting the questionnaire, respondents were then presented with the following three sections (Appendix A). A judge provided comments on an initial version of the questionnaire. These comments were used to make sure the questions would be appropriate for judges in terms of concepts used and phrasing. Respondents spent an average of 25 minutes and 25 seconds filling in the questionnaire (*SD* = 37 min and 27 s).

3.2.2.1. Knowledge and Opinion. In the first section of the questionnaire, respondents were presented with 12 statements about reasoning with evidence, the importance of excluding alternative scenarios, and looking for evidence that disconfirms the charge (e.g., it would be useful to, in addition to the case file, study the explicit arguments of the defence prior to trial; it is important that the police has investigated alternative scenarios. That should be visible in the case file). The aim of that section was to determine how judges view the role of falsification, and whether they understand its importance in light of deciding whether or not a hypothesis is true. For these statements, judges were asked whether they agreed or disagreed. Respondents were also given the chance to comment on the statements generally at the end of the section. In that way,

respondents who felt strongly about any of the statements could express their opinions in more detail.

3.2.2.2. Application Section. In the second section, respondents were presented with three short tasks that tested their focus on confirming evidence.

3.2.2.2.1. Selection task. The first of these was a variation on the Wason card selection task (Wason & Shapiro, 1971). In the Wason selection task, participants are given a conditional sentence (e.g., if there is a vowel on one side of any card, then there is an even number on its other side) and four cards. For the above example, the four cards could be the following cards: E, K, 4, 7. Participants are then asked to name those cards, and only those cards, which need to be turned over in order to determine whether the rule is true or false. In Wason's study, most participants chose 'E' and '4', or only 'E' (Wason & Johnson-Laird, 1972). The correct answer requires the card with 'E' and the card with '7' to be turned over. If there is an odd number on the other side of the 'E' card the rule has been falsified. If there is a vowel on the other side of the '7'card, the rule would also be falsified. Turning over the 'K' or '4' card would not provide any relevant information (Cosmides, 1989; Wason & Johnson-Laird, 1972). As the Wason selection task is used in some of the optional courses for judges, the task was adapted slightly in its content, but required the same logic to be solved. The conditional rule respondents received was: if a plant has round leaves, it is in a blue pot. The four options were: A plant with round leaves, a plant with pointy leaves, a plant in a blue pot, and a plant in a red pot. Following the logic of the selection task, the correct answer would be to choose the plant with the round leaves and the plant in the red pot.

3.2.2.2.2. Testing Hypotheses Task. Respondents were given a second task in which a situation with a water leak was described (Baron, 2000). A hypothesis was provided, and three possible questions were presented. The hypothesis was that the friend's son who went upstairs spilled water in the bathroom, causing the leak. One of the questions could be used to confirm the hypothesis (Is the bathroom above the kitchen?), whereas the other question could be used to

falsify the initial hypothesis (Was there a leak before the boy went upstairs?). The third question contains an alternative explanation for the leak (Is it raining?). The respondents were then asked which question they would like to see answered in order to test the hypothesis. For one of the questions, a positive answer would be expected if the hypothesis were true. For another, a negative answer would be expected. That question would be the most informative. The third question introduces an alternative scenario for the cause of the leak. That task tests the positive-test strategy, also known as the congruence bias (Nickerson, 1998; Baron et al., 1988; Baron, 2000), whereby a question that is likely to receive a positive response if the hypothesis is true, is overvalued. According to Baron (2000), subjects prefer questions to which the answer would be yes if their hypothesis was true, even if those are not the best question. The bias is thought to be reduced when alternative hypotheses are presented (Baron et al., 1988).

3.2.2.2.3. Case. In the final task, respondents were presented with a short description of an arson case. They were told that there were four witnesses, all of whom had a clear view of the perpetrator. Only one of the four witnesses identified the suspect when presented in a line-up. The other three witnesses said the perpetrator was not present in the line-up, one of which said she was sure of this. It was explicitly stated that all identification procedures were properly conducted. Respondents were then asked about their perception of the suspect's guilt based of the given evidence by choosing between a guilty and innocent response option. They were then asked to explain their decision in an open-ended answer format. Logically, the three non-identifications should be used as falsification for the hypothesis that the suspect is guilty. As argued by Sauerland and colleagues (2016), a properly conducted identification procedure with a witness who had a good view of the suspect, which results in a non-identification can be considered supportive of the suspect's innocence. However, the diagnostic value of non-identifications is often undervalued (Wells & Lindsay, 1980). The evidential value in support of the suspect's guilt of the identification decreases if two out of three witnesses who viewed the perpetrator

under similar circumstances do not identify the suspect in a properly conducted identification procedure (Sauerland et al., 2016). In the case at hand, three out of the four witnesses who had good viewing conditions did not identify the suspect in a properly conducted line-up, thus supporting the innocence of the suspect.

3.2.2.3. Experience section. In the third section, respondents were asked to provide open-ended answers to six questions about how they use falsification in practice. They were first asked to define falsification. That served as a control to ensure they understood the concept as intended. They were then asked how they attempt to falsify the charge in practice, and to provide an example from practice. Similarly, they were also asked how they attempt to verify the charge in practice, and to provide an example. The next question was whether they think it is important to falsify the charge, and to provide argumentation for their answer. They were also asked what they think is generally missing from case files in order to be able to falsify the charge. Lastly, they were asked to what extent they investigate things themselves.

3.2.3. Analysis

For the first section, for each of the statements, either the "agree" or "disagree" answer was in line with what would be expected when using logical reasoning and engaging in attempts at falsification. The percentage of respondents who answered 'agree' or 'disagree' was calculated.

For the second section, the percentages of what responses were chosen were calculated. For the task concerning identifications, we calculated how many judges thought that the suspect was guilty and examined the reasons they gave for their decision.

Responses to the third section were used to gain an impression of how judges apply falsification and verification in practice and what their experiences are with doing so. The responses are discussed broadly to give a general review of judges' approaches and experiences. The responses were also used to try to answer the question of why judges consider falsification to be important and

what they feel is missing in the case file in order to be able to attempt falsification. This final section will be supplemented with information provided by judges during the interviews. Due to the limited sample size and the variations in these responses, the qualitative data was considered insufficiently precise to provide exact measures of what participants included in their responses.

3.3. Results

3.3.1. Statements

Respondents' responses to the 12 statements are shown in Table 1, along with the percentage of respondents who agreed or disagreed with each of the statements. Respondents were unanimous about some statements, such as that a miscarriage of justice is not only caused by problems with the evidence (e.g., a false confession or wrongful identification). Furthermore, judges were also unanimous about the fact that the rest of the evidence should still be critically considered when the suspect has confessed. For other statements, responses showed a lack of consensus. For instance, 26.3% of respondents agreed that evidence for the suspect's innocence does not have to be explained if there is sufficient convincing evidence for the suspect's guilt. 15.8% also said that if there is sufficient evidence for guilt, they do not doubt their conviction despite evidence that supports the innocence of the suspect. Based on the responses, it appears most respondents understand the importance of falsification, although some answers also show that not all respondents understand the value of attempting falsification to the same extent.

Table 1

Responses to the Statements About Falsification and Alternative Scenarios (N = 36)

Statement	Agree (%)	Disagree (%)
When there is convincing evidence for the guilt of the suspect, enough investigation has been done by the police. Excluding other scenarios is not necessary. (D)	0.0	100.0
Miscarriages of justice are only caused by the evidence (e.g., wrongful identification, false confession). (D)	0.0	100.0
If a suspect has confessed, the rest of the evidence should still be critically considered, despite article 359 CCP. (A)	100.0	0.0
The evidence in the reasoned judgement that support the proven facts is the only evidence that was considered. (D)	10.5	89.5
It would be useful to, in addition to the case file, study the explicit arguments of the defence prior to trial. (A)	89.5	10.5
When reading a case file, a weighed analysis has to be made of the evidence that contradicts the charge. (A)	84.2	15.8
If there is sufficient evidence that supports the guilt of the suspect, I don't doubt my conviction, despite evidence that supports the innocence of the suspect. (D)	15.8	84.2
It is important that the police has investigated alternative scenarios. That should be visible in the case file. (A)	84.2	15.8
The testing of alternative scenarios (which are apparent in the case file or proposed by the defence) is up to the judge.(A)	84.2	15.8
The aim of the police investigation is to collect information so that a correct legal decision about the question of evidence can be made. (A)	73.7	26.3
Evidence for the suspect's innocence does not have to be explained as long as there is sufficient convincing evidence for the guilt of the suspect. (D)	26.3	73.7
There would be added value to also make the consideration of evidence that contradicts the charge visible in the reasoned decision. (A)	73.7	26.3

Note: *Letters in parentheses indicate the answers that would be desirable based on the principle of falsification and the existing literature (A = agree, D = disagree).*

3.3.2 Section 2: Tasks

3.3.2.1. Selection Task. In the first task, respondents (n = 27) were given a variation on the Wason selection task. They were instructed to choose as few options as possible. The different combinations of choices are shown in Table 2. The majority of respondents chose to only investigate the option that could confirm the rule. However, almost a quarter of the respondents also gave the correct answer (22.2%).

Table 2

Combinations of Options Chosen by Respondents. (N=27)

Options chosen	Percentage chosen
Round leaves only (Confirming)	40.7
Red pot only (Falsifying)	3.7
Blue pot only (Uninformative)	11.1
Pointy leaves and blue pot (Uninformative + uninformative)	3.7
Round leaves and red pot (Confirming + falsifying)	22.2
Round leaves and blue pot (Confirming + uninformative)	7.4
Red pot and pointy leaves (Falsifying + uninformative)	7.4
All options selected	3.7

3.3.2.2. Hypothesis Testing Task. In the second task, respondents were given a story with a hypothesis about what happened, and three questions they could ask. Two of the questions focused on the existing hypothesis. The majority of respondents chose to ask whether it was already leaking before the boy went upstairs (68.4%). That question was the most informative (Baron, 2000). The confirming question whether the bathroom was located above the kitchen was chosen by 26.3% of respondents. Only 5.3% of respondents chose to ask whether it was raining, thereby testing another explanation for the water leaking into the kitchen. Participants therefore did not show a preference for the question that

would result in a positive answer if the hypothesis were true, with the majority selecting the question that would be most informative to test the hypothesis.

3.3.2.3. Case. When presented with the description of the arson case, and the outcomes of the four identification procedures, participants were asked whether they would convict or acquit the suspect, and why. The vast majority of participants chose to acquit the suspect (74.1%). In their reasoning, they mostly referred to the three non-identifications, or not having enough evidence. Of the respondents who chose to convict the suspect (15.8%), one respondent also said that they did not have enough information. One respondent who would convict the suspect stated that a lot of people are bad at recognising people. They mentioned the witness who identified the suspect as the perpetrator, saying that she had a good view, and that the suspect was connected to the property. Another respondent who chose to convict the suspect also referred to the identification, as well as the suspect's prior conviction. None of the respondents who chose to convict the suspect mentioned the non-identifications. Two respondents (7.4%) did not make a decision for this task.

3.3.3. Section 3: Open Questions

The open questions that were presented to the respondents in the survey were subsequently also the basis for the semi-structured interviews with three judges. In this section, the answers from the survey respondents will be combined with the information provided by the interviewees.

3.3.3.1. Understanding and Application of Falsification. Respondents were first asked what falsification is according to them. The majority of judges answered this question quite well, showing a good understanding of the concept of falsification. Their answers included things such as "investigating whether a statement could be untrue", and "trying to disprove a claim". Several answers also included showing, or making it likely, that a theory or statement is incorrect. One answer mentioned the theory of Popper, stating in the judicial world, it means going over contradictory, alternative evidence or scenarios.

When asked how they try to falsify the charge, a lot of the respondents mentioned the consideration of alternative explanations for the evidence, or alternative scenarios about what happened, with most emphasising the exclusion of alternative scenarios. Some also mentioned looking for indications in the case file that the charge against the defendant might be wrong, or weighing evidential value of the incriminating and the exonerating evidence. Several judges also gave responses in line with "looking at how the evidence fits with the theory that the suspect is innocent", and "looking critically at the evidence for facts and circumstances that are exonerating for the suspect".

One judge explained that they try to falsify the charge by testing the charges in the indictment against the full range of evidence in the dossier. If there is insufficient evidence, or an alternative scenario that cannot be excluded, the charge is not necessarily falsified, but you would have to acquit, due to the possibility that the charge is wrong. In sum, it appears that the judges' understanding of how to effectively falsify differs. Whereas all respondents seemed to understand the importance of alternative scenarios, they had various levels of understanding of how alternative scenarios can be used to attempt falsification.

During the interviews, it also became apparent that there is no clear expectation among judges of how they should attempt falsification. For instance, some respondents expressed that it was unclear whether judges should do more than look for evidence in the case file that contradicts the charge and what the weight of the evidence is. Whether judges should go further, and whether or not the type of case or the approach of the defence should affect how far you go, all remain unclear, according to some of the respondents. One of the interviewees also mentioned that if the file contains elements which do not make sense with the charge, it can create sufficient doubt for an acquittal. Another interviewee made it clear that when reading the case file she looks for evidence both for and against the charge against the suspect. She would then try to clarify that evidence during trial. However, she also mentioned that she does not specifically

look for alternative scenarios, but when there are alternative scenarios that are realistic and apparent from the case file, it can affect whether or not the judge is convinced.

Respondents were asked if they could give an example of how they attempt falsification in practice. The answers were in line with the varying levels of understanding observed throughout the survey responses. Several judges gave examples of the suspect having an alibi to falsify the charges. Others also mentioned investigating whether there is another explanation for the suspect's DNA being found at the crime scene. Another respondent also gave the example of an abuse case with two reports. They would then check whether the reports support each other on crucial points, whether they are consistent and detailed, and whether there was any influence between the reporters. The respondent did not specify what points they would want to check. Although that response demonstrates a very critical approach towards evidence, it does not necessarily result in falsification attempts. Another judge also responded that it should be investigated whether an unknown perpetrator could have committed the crime instead of the suspect, or whether the victim could have died due to a cause other than the actions of the suspect. If these investigations focus on trying to find evidence for these scenarios rather than on trying to disprove them, that would be a good example of attempted falsification. Therefore, the answers, again, show a range in the understanding of falsification, and the manner in which different judges attempt falsification in practice.

3.3.3.2. Understanding and Application of Verification. Respondents mostly referred to the evidence in the case file when asked how they try to verify the charge. They would check the charge based on this evidence, or they would determine how the evidence in the case file fits with the theory that the suspect is guilty, as well as the evidence that the suspect is innocent. Although some respondents here only wrote about the evidence supporting the charge or about finding sufficient legal evidence, some also mentioned basing verification on both the incriminating and exonerating evidence in the case file. One respondent

also explained the difficulty of determining how convincing the evidence should be considered to be. The respondent further points out that "besides the decision maker's own belief, the collective conviction that arises (from multiple judges)" can also help with this determination. The general impression based on these answers is that verification is done based on the case file, and that some judges clearly understand verification and falsification to be processes that are closely associated.

When asked to give an example of how they verify a charge in practice, several respondents mentioned in more general terms that it was a daily practice for them. Others also referred to the minimum requirements for evidence. For instance, having CCTV footage of someone stealing beer, and finding the beer in that person's pocket.

3.3.3.3 Importance of Falsification. In response to whether they believe it is important to falsify, all except one of the 17 respondents answered positively. The single respondent who answered negatively explained that the statement lacked nuance and that in some cases, there is no added value in trying falsification. In cases where there is added value, they would consider it important. Unfortunately, they did not specify what kind of cases this would include. Another respondent explained that it becomes more important with more indications for alternative scenarios, and that those are often lacking. One respondent explained that falsification is important when the suspect denies or gives another version of events, but is less important when the suspect provides a supported confession, whereas another said it is important to "prevent tunnel vision, even when a suspect confesses". Many of the answers included that falsification can protect against tunnel vision, with one explaining that "only looking for supporting evidence can lead to not seeing evidence that could disconfirm the statement, even though the latter is more decisive than more supporting evidence". Several respondents also explained that having sufficient legal evidence does not necessarily mean the suspect is guilty. Overall, respondents seemed to think falsification was important to prevent tunnel

vision, but there was again variation in the types of cases this was considered a necessity for. According to one of the interviewees, it is now generally accepted that judges should try to avoid tunnel vision, and be as open-minded as possible. One interviewee said that a good judge will usually be forced by the different parties to do so, whereas another interviewee mentioned that it is more or less expected to see what exonerating evidence there is. One of the interviewees remarked that working with alternative scenarios and the critical thinking is part of the criminal procedure, and that he feels falsification is part of the procedure.

Survey respondents were also asked if they felt they were able to falsify sufficiently in every case. Only one respondent answered yes. When asked what is important to falsify, there was again a lot of emphasis placed on the alternative scenarios – whether they were investigated during the police investigation, or whether plausible alternative scenarios that cannot be excluded can be investigated further. Others also felt the core of the accusation was important for falsification.

The respondents who answered 'no' were subsequently asked what they felt was missing in the case file to be able to falsify. Respondents again referred to alternative scenarios which were investigated by the police, or insufficiently supported by the defence lawyer. One respondent specifically mentions that if a suspect does not make a statement, or denies, an alternative scenario is not always looked for or investigated, and the reliability of, for example, witnesses is not checked. The respondent explained that some investigative measures are then not used because the perpetrator is thought to be in custody, and that bad handling of the crime scene may make it impossible to conduct further investigations at a later stage. Another respondent mentioned that the police investigation is limited to finding the evidence for the scenario that is most obvious, and that that is inherent to the available resources. Several respondents also answered that factual information that can be checked is missing from the case file. Another respondent specifically stated exonerating facts and

circumstances were lacking from the case file. One respondent also pointed out that the defence sometimes only brings up an alternative explanation during trial. Lastly, one respondent also stated a lack of time as an issue, and wrote that the case file does not provide the information that is necessary to attempt falsification. The lack of time judges have to prepare or study the case file was also given during the interviews as a contributing factor to not being able to attempt falsification.

3.3.3.4. Judge's Role. During the interviews, the respondents also said that they were mostly dependent on what the police had done or alternative explanations that had been brought up by the defence. If the defence proposes an alternative explanation, it will need to be determined how plausible that explanation is and whether further investigation is needed. Keeping capacity in mind, further investigation will often not happen. One of the respondents also mentioned that the judge is reliant on the file, even though it may not include relevant information which was deemed unnecessary in an earlier stage of the investigation.

As the penultimate question, respondents were asked to what extent they actively research cases themselves, for instance by investigating alternative scenarios, or gathering evidence either during or outside of trial. The answers again showed a big range of different approaches. Whereas some respondents said they do not do this, others felt it was a continuing important part of the job, as it goes with determining the truth. Several respondents said they will check alternative scenarios based on the evidence in the case file, but not necessarily do any other investigation themselves. Some judges explained that they would for instance use Google street view to get a better impression of the situation. One judge acknowledged that while reading a scientific article can help you, bringing in this knowledge will require an expert to be called in. Most of the further research by judges is conducted by questioning of witnesses, experts, or the suspect. One respondent explained that if it is unclear how a piece of evidence should be interpreted or evidence is missing because an important

aspect has not been investigated, then further investigation should be done. The responses also made it clear that it will depend on the type of case and how the police investigation was completed. Some of the respondents and interviewees explained that they consider it the defence's task to present the alternative scenarios and request further investigation, whereas others struggled with in how far the judge can focus on alternative scenarios if these are not proposed by the defence. In sum, the extent to which judges feel the need or responsibility to do further investigation again varied greatly between the different respondents.

During the interviews, the concept of 'active judge' also came up several times. There it was again apparent that there is a lot of variety in how that role is interpreted, for instance in the extent to which judges feel they should, or can, do further investigation. Whereas some judges feel doing further investigation could taint their impartiality, others feel it is a part of their job. It could also lead to more acquittals: judges might need more evidence to be convinced, but are hesitant to request the further investigation to get that evidence, resulting in insufficient evidence to reach conviction. There are also different ways in which the judge can be 'active', for instance by requesting further investigation, but also by critically questioning the expert witness. The benefit of an 'active judge' is threatened by being biased or prejudiced. Why one judge might choose to be more active than another judge could be due to the specific case, but could also be due to the individual differences. One of the judges also made clear that the active judge is very important during the pre-trial review. Overall, it appears there are several unanswered questions for judges themselves when it comes to how active they should be, and how falsification fits in with that task.

In the final question, respondents were asked what the most common reason was for them to want to request further investigation. Several answers related to not being convinced about the events as described in the charge, or wanting to avoid making a decision that does not correspond to what happened. Some respondents also gave insufficient police investigation as a reason for wanting to do further research. One respondent stated that they would do further research

to prevent too quickly accepting an obvious scenario. Another said that a witness who denied the charge would be a reason to do further investigation. Overall, the reason for doing further investigation can, in some cases, be related to engaging in falsification or to excluding alternative scenarios.

With regards to scenario consideration, it is not clear what is expected of the judge, which results in judges making an individual decision. There are minimal rules about evidence, and it is relatively easy to come to a construction that is legally sufficient. One of the interviewees also said that she can get lots of evidence that supports the charge from the case file, but that does not necessarily mean she would be convinced. Often the problem is not in meeting the minimum requirement, but in the weighing of the evidence and how diagnostic the evidence is for guilt or innocence. One of the judges specified during the interview that there may be enough evidence to convict, but if she is not convinced she will acquit, and similarly, if there is not enough evidence, she must also acquit, despite being convinced. That illustrates that the legally required evidence does not necessarily mean a judge will be convinced.

3.3.3.5. Other relevant remarks made during the interviews.

3.3.3.5.1 Clerk. During the interviews, it became clear that the clerk plays a large role in the preparation of the case file, which could affect the consideration of evidence. The clerk makes an overview of the evidence in the case file, which may focus mainly on the incriminating evidence. One of the interviewees said that the overview clearly includes both exonerating and incriminating evidence. Due to the limited time judges have, the overview can crucially affect how judges read the case file. However, there were also differences between the interviewees in how they read a case file. Whereas one of the interviewees felt it was unusual to simply read the case file from the beginning, another interviewee said that was how she reads a case file.

3.3.3.5.2. Reasoned Decision. When asked whether it would be useful to also include alternative scenarios in the reasoned decision, one of the interviewees explained the existing requirements as they experience them. There is a

requirement to, using the evidence, indicate what evidence confirmed for the charge, but when an alternative scenario is raised, what you have to do depends on the plausibility of the scenario. There are three categories according to this interviewee: those where you need to use evidence to exclude the scenario, those which can be argued away because they are not very plausible and those that can simply be ignored. There is also a need to respond to the defence's substantiated arguments about the reliability of evidence, which we interpret as a need to pay attention to the validity and consistency of the evidence. In that sense, there has been a lot of progress, but there are not many requirements for the reasoned decision. If you know that you are going to have to explain why you think something is reliable at a later stage, it can influence how you look at the evidence. However, the interviewee did not think that the reasoned decision is already considered when reading the case file.

Another interviewee also said that things in the file that might not make sense will be discussed, but will not necessarily be included in the written judgement. He argued that it can differ between judges, as they are not required to include everything because the judge, in principle, selects the evidence. However, the written decision should be logical, and if important things are left out, he expected it to cause problems for the court of appeal or the Supreme Court. Another interviewee also said that evidence that contradicts the charge would not be in the reasoned decision if the charge is considered proven. According to her, the reasoned decision includes how the judge believes the event happened, and should include why they think it is proven, not why they think it would not be proven.

3.4. Discussion

The results of the current survey offer novel insight into the application of falsification in practice, and moreover, into judges' perceived need and opportunity to use falsification in the way that we would theoretically consider it a necessary step in logical reasoning. The overall impression gained suggests

that although the majority of judges appreciate the need for falsification, there is quite a wide range in interpretation of how one should attempt to use falsification when deciding whether or not a suspect is guilty. In addition to the individual differences between judges, it also appeared judges struggle to use falsification in practice due to other reasons, such as the construction of the case file or time pressure.

For the various statements, the majority of judges demonstrated having knowledge and opinions in line with the added value of falsification. More than a quarter of the judges we surveyed felt that evidence for innocence does not have to be explained as long as there is sufficient evidence for guilt. As was mentioned during the interviews, it is quite easy to reach the minimum requirement of evidence to prove the charge. Similarly, it is also concerning that 10% of the judges agreed that the only evidence considered is the evidence that is included in the reasoned decision as supporting the suspect's guilt. That would suggest that potentially exonerating evidence was not considered. Such a focus on incriminating evidence could indicate that confirmation bias could affect the decision-making process. Furthermore, the requirements for the reasoned decision appear to also increase the risk of excessively focusing on evidence that supports the charges. It should be noted that several judges commented on a lack of nuance in the statements. A few judges expanded on their point, for instance by saying that they would agree with some of the statements in an ideal world, but that in practice, it is not so absolute. If possible, it would therefore be interesting to extend the current survey by giving judges a better platform for their opinions about the statements, for instance in the form of an interview or a focus group.

Based on the tasks in the second section of the survey, the respondents again showed a relatively good ability to consider evidence that was not confirming for their hypothesis. In the variant of the Wason selection task, 40% of respondents chose only the confirming option. However, 22% of respondents also correctly chose both the confirming and the falsifying option. Typically, fewer than 10% of

the participants give the correct answer to the Wason selection task (Rachlinski, 2012). In previous research, Rachlinski (2012) also described that out of Dutch judges who completed a less abstract variation on the Wason selection task, 32 % chose the correct answer, thereby outperforming the U.S. judges who were tested. Therefore, although the respondents' performance may at first glance not seem positive, the result is relatively good considering the task.

For the hypotheses testing task, respondents again performed well. Although a significant minority chose the question which contained an alternative explanation, only a quarter of respondents chose the question which would have had a positive answer had the hypothesis been true. The majority of respondents chose to ask the question which has the lowest chance of receiving a positive answer if the hypothesis were true, and is subsequently the most informative (Baron, 2000). Respondents therefore did not seem to show the positive test strategy. According to Kahnemann (2011), the positive test strategy tested by this task is akin to confirmation bias, namely by constituting a search for confirming evidence. Therefore, the results of the current task suggest that the respondents were not excessively focused on finding confirming evidence.

In the final task, the respondents were asked to decide whether or not they would convict a suspect, based on the evidence that was given, which included one identification and four non-identifications. The majority of respondents chose not to convict the suspect, with many stating that the one identification was not sufficient and that the only witness who identified the suspect could not connect the suspect to the crime. Many also gave the three non-identifications as a reason for their decision. However, some respondents chose to convict the suspect. In their explanation of their choice, none of these respondents referred to the non-identifications. The variability of the decisions made for this task also reflected the range of approaches to exonerating evidence that were explained by the respondents in the open questions in the third section of the survey.

In the open questions, the respondents generally showed a good understanding of what falsification is when asked how they would define

the term. However, when asked how they attempt falsification in practice, respondents mostly explained that they would try to exclude alternative scenarios. Whereas the exclusion of alternative scenarios is an important aspect of trying to determine what happened, it is also an indirect form of finding support for the main scenario. When someone tries to disprove an alternative scenario, he is indirectly trying to find support for the initial scenario. In order to use alternative scenarios for falsification, one should try to both confirm and disconfirm the alternative scenarios, as well as the main scenario. As was also demonstrated in a Chapter 2, alternative scenarios did not seem to increase the attempted falsification, as neither support for the alternative scenario, nor contradictions for the main scenario, appeared to be seriously considered.

The concept of an active judge was discussed during the interviews, and could also be found in the answers given to the open questions. The idea of an active judge, which is also an element of the inquisitorial system, is commonly accepted amongst Dutch judges (Cleiren & Dubelaar, 2014), although there is a wide variety in how the concept is interpreted, and how this role is fulfilled. The concept can be closely related to the judge's perceived need to falsify. For instance, although the majority of judges felt it was their job to consider alternative scenarios, there appeared to be no consensus on whether these scenarios should be considered if apparent in the case file, or only if they were raised by the defence. Furthermore, the idea of requesting further investigation is also closely related to both the concept of an active judge and falsification. For instance, active can mean critically questioning a witness, but can also mean requesting further investigation of an alternative scenario. It could also mean that the active judge would request further investigation of the alternative scenario whereas the less active judge would conclude that the scenario cannot be excluded and would acquit the suspect. The varied interpretation of the role has previously been acknowledged in the literature (Ferdinandusse, 2018; De Weerd, 2013). Ferdinandusse (2018) also expressed that a more explicit agreement on what the active role of the judge should be is desirable, as the difference

in how active a judge is could influence whether the same suspect would be acquitted or convicted.

From the open questions and the interviews, it also became clear that, despite judges' good intentions, the majority of respondents said they did not feel able to falsify in every case file. Although it could be argued that the need for falsification also varies between cases, it should nevertheless be possible, at least to some extent in most cases. The lack of opportunity could be due to several issues. It could for instance be due to the construction of the case file, where information may have been left out because it was deemed irrelevant at an earlier stage. Furthermore, it could also be due to time pressure, which is a common problem among judges in The Netherlands at the moment. There have been several reports of judges having a workload that is too high due to a number of issues, including a lack of funding and staff. The threats that these issues pose for the quality of the legal system have been described by both the Dutch Association for the Administration of Justice (2018), and Tegenlicht, a group of concerned judges (2018).

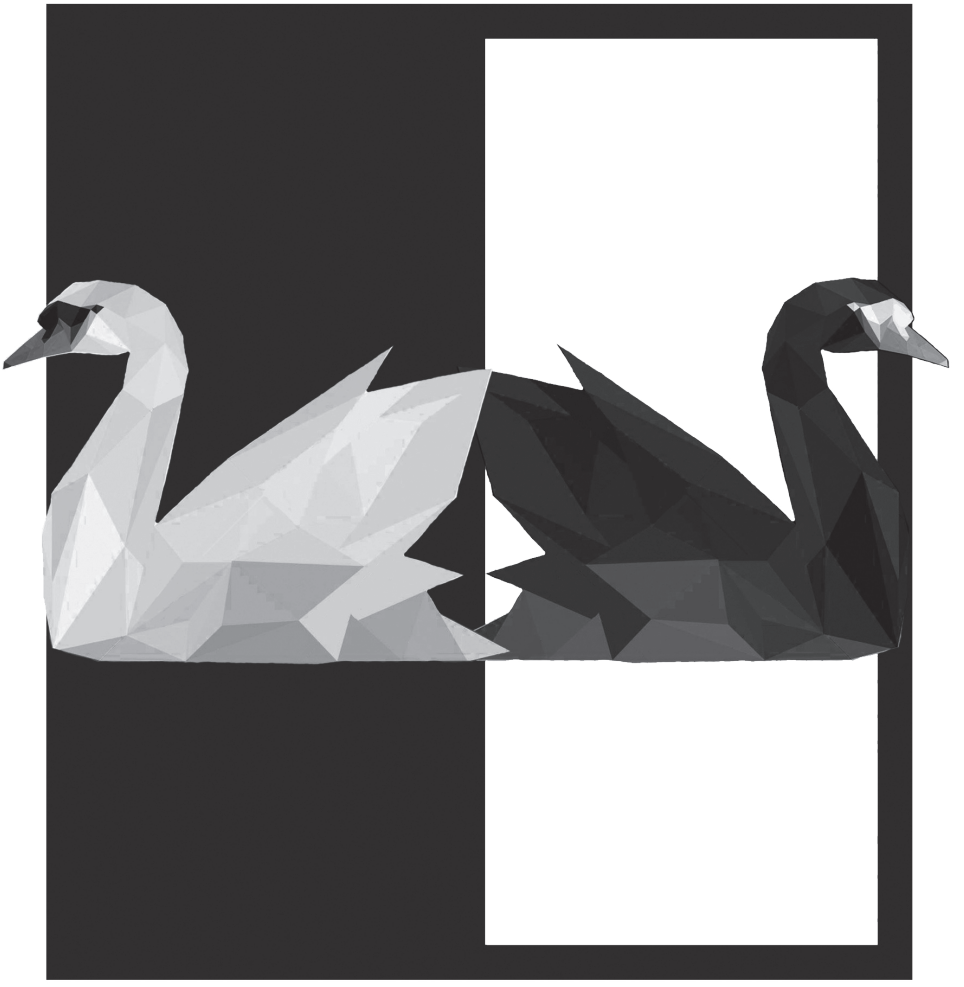
Unfortunately, the current survey did not specifically investigate the obstacles judges experience when trying to use falsification. However, a number of areas that could be researched further in order to elucidate ways in which falsification could be facilitated became clear. For instance, falsification could be encouraged through emphasising and clarifying the role of the active judge. It would be interesting to gain insight into the interpretation of that role among a bigger group of judges.

Another interesting area for further research would be to look at the role of clerks. It became clear that their preparation of the overview of the case file can be helpful for judges when they first start reading a case file. However, it also seems like there is no requirement for the clerks to include exonerating evidence in the overview. That is unsurprising when considering that the main question to be answered by judges is whether it is proven that the act in question was committed by the suspect (Art 350 CCP), which can only be decided in case

the judge has been convinced on the basis of legal evidence (Art 338 CCP). As mentioned during the interviews, a convincing construction can quite easily be achieved. Hence, when trying to determine what happened, it could be helpful to also include evidence that contradicts the charge. It could offer a safeguard against becoming convinced too easily, as it could make it easier to consider the evidence for and against multiple scenarios, thereby facilitating attempts at falsification of the potential scenarios.

One limitation to keep in mind when considering the results of the current survey is the limited sample size, as well as the fact that the sample may not have been representative of judges in The Netherlands overall. The judges took part in the survey voluntarily, and therefore may have been judges with a special interest in the topic of the survey. In order to get a more representative sample of judges' perception and explanation, many more judges would need to be surveyed or interviewed. However, despite several attempts to recruit more respondents or interviewees, we were unable to increase our sample size.

The findings of the current survey extended the findings of Van Veldhuizen et al. (2019), namely that judges value the importance of falsification and of ruling out alternative scenarios. The findings further indicate that the judges performed relatively well on the tasks testing for a bias in favour of confirming evidence. However, the open questions revealed that judges may be excessively focused on excluding alternative scenarios. Furthermore, there appears to be a need for clarification of what is expected of an active judge. As judges also explained that they struggle to use falsification in practice due to a number of issues, methods in which falsification can be facilitated should be researched, especially considering its importance as acknowledged by judges.



CHAPTER 4

LAW AND ORDER EFFECTS: ON COGNITIVE DISSONANCE AND BELIEF PERSEVERANCE

An adapted version of this chapter has been accepted for publication as:

Maegherman, E., Ask, K., Horselenberg, R., & Van Koppen, P. J. (2020) Law and order effects: On cognitive dissonance and belief perseverance. *Psychiatry, Psychology, and Law.*

Abstract

Order of evidence presentation affects the evaluation and the integration of evidence in mock criminal cases. In this study, we aimed to determine whether the order in which incriminating and exonerating evidence is presented influences cognitive dissonance, and subsequent display of confirmation bias. Law students (N = 407) were presented with a murder case vignette, followed by incriminating and exonerating evidence in various orders. Contrary to a predicted primacy effect (i.e., early evidence being most influential), a recency effect (i.e., late evidence being most influential) was observed in ratings of likelihood of the suspect's guilt. The cognitive dissonance ratings and conviction rates were not affected by the order of evidence presentation. The effects of evidence presentation order may be limited to specific aspects of legal decisions. However, there is a need to replicate the results using procedures and samples that are more representative of real-life criminal law trials.

4.1. Introduction

In criminal law trials, the main question to be answered by the judge or jury is whether it is proven beyond reasonable doubt that the defendant committed the crime he or she is accused of. The judge or jury must do so based on the evidence that has been presented to them by the prosecution and the defence. That process will usually require the judge or jury to draw inferences from the evidence, and to combine these to reach a final decision on guilt. The reasoning process required to reach a decision on guilt ultimately relies on human cognition, and is therefore likely to be vulnerable to cognitive biases. Such biases can affect how information is processed, including how it is evaluated and integrated (Charman et al., 2019). One form of bias that has been argued to be particularly relevant to the legal system is confirmation bias (Findley & Scott, 2006). In this study, we aimed to determine whether the order in which evidence is presented influences mock legal decision-makers' cognitive dissonance and the subsequent presence of confirmation bias in the context of criminal proceedings.

4.1.1. Biased Reasoning

Confirmation bias refers to the human tendency to seek and interpret evidence in ways that favour an existing belief, thereby influencing reasoning and decision-making. That influence can have a negative impact. It has been argued to be one of the most problematic aspects of human reasoning (Nickerson, 1988). The existence of confirmation bias has been demonstrated in various areas of criminal investigations, ranging from interviews to forensic investigation (e.g., Kassin et al., 2003; Dror et al., 2006; Ask & Granhag, 2005).

In order for confirmation bias to influence reasoning, one first needs to hold a belief. Therefore, confirmation bias is closely related to belief perseverance, a basic human tendency to adhere to accepted theories or explanations even when faced with discrediting information (Burke, 2007). Multiple researchers have shown that changing people's initial beliefs by confronting them with contradicting evidence can be difficult to achieve. In a classic study, Lord

et al. (1979) presented participants with information either confirming or disconfirming their view on capital punishment. Participants found the evidence confirming their initial belief to be more convincing, and there was a lack of belief adjustment in response to contradicting information. In a more recent study, Green and Donahue (2011) presented participants with a story about a young heroin addict. Participants were then told that the story was inaccurate, due to either accidental error or intentional deception. Although those in the intentional deception condition were motivated to change their belief, participants in both conditions maintained beliefs about items specific to the story (e.g., failure of social programmes to assist young people). Furthermore, Burke (2007) describes how, in many exoneration cases, prosecutors have maintained that the exonerated individual is guilty despite overwhelming evidence to the contrary. In short, belief perseverance is a well-documented issue, for which effective solutions are currently lacking.

Belief perseverance has also been researched in the applied setting of legal proceedings. Schünemann (1983, as cited in Schünemann & Bandilla, 1989) offered valuable insights by investigating belief perseverance in judges. In his study, judges who had been given more incriminating information prior to trial were more likely to convict the defendant than judges who were given the same case file, but less incriminating prior information. Therefore, judges also appear to be prone to belief perseverance despite the need for impartiality.

In the Netherlands, the case file judges read before the trial proper is likely to consist mainly of incriminating information (Crombag, 2017). That case file, or a summary of the case file, is likely to form the basis of the judge's initial belief. Although the case file is mostly incriminating, it is also likely to contain some information that is contradictory to the suspect being guilty. Such information will often be emphasised by the defence. The judge would then have to process arguments supporting two different beliefs. Experimental psychological research has shown that the consideration of conflicting beliefs can create an uneasy feeling, known as cognitive dissonance (Festinger, 1957). One way in

which people try to escape cognitive dissonance is to adopt, and adhere to, one of the beliefs, while refuting or downplaying the other (Jonas, Traut-Mattausch, Frey, Greenberg, 2008).

Cognitive dissonance theory has been researched in various experimental studies. Harmon-Jones and Harmon-Jones (2007) reviewed the development of Festinger's theory and concluded that cognitive dissonance is mainly due to inconsistency rather than the possibility of aversive consequences due to the chosen alternative. Legal psychological research indicates that cognitive dissonance may arise in the presence of belief-inconsistent criminal evidence (Ask et al., 2011). Participants who received evidence inconsistent with their initial belief about a suspect's guilt reported experiencing more dissonance than those who received consistent evidence. They also found that those who experienced the strongest dissonance were the most reluctant to adjust their judgement of guilt in line with the contradicting evidence. Moreover, participants who received contradicting witness evidence reported experiencing stronger dissonance than those who received contradicting DNA evidence. According to Ask et al. (2011), that difference can be explained by the fact that witness evidence has higher elasticity (i.e., room for subjective interpretation) than DNA evidence. Thus, participants had little choice but to change their belief in line with the compelling DNA evidence and, as the conflict between beliefs was resolved, these participants experienced less dissonance. In contrast, the contradicting witness evidence evoked stronger dissonance as it created conflicting beliefs, none of which was clearly more compelling than the other (Ask et al., 2008). In sum, Ask et al. (2011) found support for the fact that the cognitive dissonance experienced is related to the evaluation of criminal evidence. They also found support for a mechanism whereby cognitive dissonance can be reduced—ignoring contradicting evidence.

The preference for supporting information in response to cognitive dissonance is akin to confirmation bias (Jonas et al., 2001). Therefore, manipulating belief perseverance and cognitive dissonance may affect the presence of confirmation

bias. Whether cognitive dissonance is affected by the presentation of multiple pieces of evidence, varying in consistency with an existing belief, has not yet been researched.

4.1.2. Order Effects

As shown above, cognitive dissonance is associated with a reluctance to adjust perceptions of guilt in line with contradicting evidence. A way to reduce such belief perseverance may be to present evidence to jurors in a way that prevents the formation of a strong prior belief and, subsequently, the development of dissonance. For instance, less cognitive dissonance may be experienced when one is, from the outset, presented with both supporting and contradictory information, rather than first receiving all the evidence in favour of one view, followed by all the evidence in favour of the other. In other words, it may be possible to reduce the development of confirmation bias by changing the order in which evidence is presented.

Effects of the order in which evidence is presented were examined by Pennington and Hastie (1988). Participants perceived evidence as stronger when it was presented in a story order rather than in the order in which the witnesses provided their statements. When the prosecution's evidence was presented as a story, a majority of participants convicted the suspect. Similarly, when the defence's evidence was presented in the story-order, a majority of participants acquitted the suspect. The totality of evidence presented to participants in these conditions was identical, suggesting that the difference in verdicts must be due to the different presentation orders. The importance of a coherent story has also been recognised in other influential theories on legal decision-making (e.g., Simon, 2004; Wagenaar et al., 1993).

In previous research, the order in which evidence is presented has been investigated mostly in relation to specific types of evidence (e.g., alibi, DNA, or witness evidence). For instance, Price and Dahl (2014), who used alibi and eyewitness evidence, found an influence of strong evidence that was presented

last on the evaluation of the suspect's guilt overall, as well as on the evaluation of individual pieces of evidence. The finding of such a *recency effect* is contrary to what would be expected in line with belief perseverance; belief perseverance can be considered a form of *primacy effect*, where the conclusion is based disproportionately on the initial evidence.

Charman and colleagues (2016) made use of DNA evidence and alibi evidence to test the influence of order of presentation. They concluded that the initial evidence had an influence on the evaluation of subsequent evidence, which can be considered supportive of a primacy effect and belief perseverance. Moreover, they concluded that the piece of evidence which is presented last does not retroactively influence evaluations of the evidence that was presented earlier. However, similarly to the finding by Price and Dahl (2014), Charman and colleagues (2016) also found that the piece of evidence which was presented last had a greater impact on the overall assessment of guilt. Charman et al. (2016) therefore suggested that *evidence evaluation* is likely influenced by an existing belief, whereas the *evidence integration* is likely to be influenced by a recency effect. As these studies on order effects looked only at the presentation of two pieces of evidence, the effect of alternating the presentation of incriminating and exonerating evidence remains unexamined.

4.1.3 The Current Study

The aims of the current study were two-fold. Firstly, we attempted to replicate the previous finding that participants maintain their initial beliefs when presented with contradicting information (i.e., belief perseverance; e.g., Green & Donahue, 2011). Secondly, we investigated whether the order in which incriminating and exonerating evidence are presented to participants affects the occurrence of cognitive dissonance and belief perseverance. The types of evidence used in this study included hair, CCTV footage, shoeprints, and witness evidence, thereby also extending previous research which used a limited number of evidence types.

Participants in this study were first presented with a summary description of a homicide case, in which a suspect had been arrested. Their cognitive dissonance was then measured, as well as their initial impression of the suspect's guilt. They then received two pieces of either incriminating evidence, exculpatory evidence, or a mixture of incriminating and exculpatory evidence. After having rated the likelihood of the suspect being guilty and their experienced cognitive dissonance, they received an additional two pieces of either exculpatory evidence, incriminating evidence, or a mixture of the two. Depending on experimental condition, the valence of the evidence received in the second set was either similar to or different from the evidence participants received in the first set. Participants then made final ratings of suspect guilt and cognitive dissonance.

First, we predicted that the order in which the evidence is presented would make a difference to participants' final guilt ratings. Specifically, participants who received incriminating evidence in the first set and exonerating evidence in the second set (*Inc/Ex*) would make higher final ratings of guilt than participants who received mixed evidence in both sets (H1). Those in the mixed conditions were not expected to strongly commit to one belief, as they had received both incriminating and exonerating evidence. However, those in the contradictory conditions had only seen either incriminating or exonerating evidence and were therefore expected to commit to guilt or innocence, respectively. Similarly, participants who received exonerating information in the first set and incriminating evidence in the second set (*Ex/Inc*) would maker lower final ratings of guilt than participants who received mixed evidence in both sets (H2). Because all the above participants had received the same pieces of evidence in total, but in different order, the predicted effects can be attributed to order effects.

Second, we expected that participants in the contradictory conditions would show belief perseverance, thus preserving their initial belief when presented with contradictory evidence. Hence, we predicted that participants in the *Inc/Ex* condition would make higher final judgements of guilt than participants in the *Ex/Inc* condition (H3).

Third, we predicted that the final judgment of cognitive dissonance would differ between conditions. Specifically, participants who received mixed evidence in both sets would report less cognitive dissonance than participants who received contradictory evidence (i.e., *Inc/Ex* and *Ex/Inc*; H4), but more cognitive dissonance than participants in the control conditions (who received uniformly incriminating or exonerating evidence across both sets; H5). Finally, we explored whether the amount of cognitive dissonance was associated with changes in guilt ratings between the first and the second set of evidence; that is, whether dissonance was associated with belief perseverance, as reported in previous research (Ask et al., 2011). As no such hypothesis was included in the preregistration, that was tested exploratorily. The preregistration for this study can be found at https://osf.io/ywajg/?view_only=a94c0cd904ed4c2cbe3bb8d82e5185b6.

4.2. Method

4.2.1. Participants

Law students were recruited through several communication platforms at Maastricht University as well as through social media and other forms of advertisement (e.g., flyers handed out at university buildings). It was decided to use law students due to their affinity with the context of the study, namely legal proceedings, as well as their familiarity with the decision that participants were required to make. Professional judges are notoriously difficult to use as participants in experimental research. As the focus of the study was decision-making by judges in a European, inquisitorial system, law students were considered to be the most appropriate sample.

Participants could win a shopping voucher worth € 10 for their participation in the study. An *a priori* power analysis, conducted using G*power 3.1 (Faul et al., 2009), showed that a total of 276 (46 per condition) participants were needed to detect a small effect size ($f = 0.10$) with 80% power at a significance level of .05. Data were collected in both Dutch and English at the same time (see below). Random allocation was used in both language versions, which resulted in some

groups receiving more participants than necessary. It was therefore decided, prior to conducting any analyses, to add participants in the remaining conditions to achieve approximately equal group sizes. Across both samples, a total of 474 participants completed the survey. In total, 67 participants were excluded from the initial sample for three reasons: Firstly, 51 participants failed to complete the initial cognitive dissonance measure. Secondly, 13 participants were removed because they had an educational background other than law. Finally, three participants completed the study in under 240 seconds. Prior to conducting any analysis, that completion time was deemed insufficient to read through the material properly, and these participants were also excluded. After exclusions, the final sample consisted of 407 participants. Participants' average age was 22.97 years (*SD* =4.24) and the majority of participants were female (62.7%). The survey was offered in both English and Dutch and participants chose their preferred language. Most participants completed the survey in English (58.7%). Ethical approval for this study was obtained from the Ethical Committee of the Faculty of Psychology and Neuroscience at Maastricht University.

4.2.2. Design

The study used a 6 (evidence order) × 3 (time) mixed factorial design, with time as within-participants factor. Participants were randomly assigned to one of the six evidence order conditions specified in Table 1. In two of the conditions, participants received a mix of incriminating and exonerating evidence at both Time 2 and Time 3. Those two conditions differed in that one of them started with incriminating evidence as the first of two pieces of evidence at Time 2 (*Mixed 1),* whereas the other condition started with exonerating evidence *(Mixed 2).* Another two conditions consisted of contradicting evidence between Time 2 and Time 3. In those conditions, participants received either only incriminating (*Inc/Ex)* or only exonerating evidence *(Ex/Inc)* at Time 2, and then the contradicting evidence at Time 3. In the last two conditions, which served as control conditions, participants received only exonerating *(Ex/Ex)* or incriminating evidence *(Inc/Inc)* at both Time

2 and Time 3. All participants were asked to rate the measures outlined below after reading only the case description (Time 1), after receiving the first set of evidence (Time 2), and after receiving the second set of evidence (Time 3).

Table 1

Representation of the Order Presentation in the Conditions.

	Evidence	
Condition	First set	Second set
Mixed evidence		
Mixed 1	Incriminating/Exonerating	Incriminating/exonerating
Mixed 2	Exonerating/Incriminating	Exonerating/Incriminating
Contradictory evidence		
Ex/Inc	Exonerating/Exonerating	Incriminating/Incriminating
Inc/Ex	Incriminating/Incriminating	Exonerating/Exonerating
Control conditions		
Ex/Ex	Exonerating/Exonerating	Exonerating/Exonerating
Inc/Inc	Incriminating/Incriminating	Incriminating/Incriminating

4.2.3. Materials

Participants completed an experimental online survey administered using the Qualtrics survey platform. The survey consisted of three components: a case vignette, evidence, and a set of measures (Appendix B).

4.2.3.1. Case Description. Participants were given a vignette describing the early stages of a homicide investigation. The case was based on material previously used by Ask et al. (2008) and was adapted for use in the current study. The information in the vignette was intended to be ambiguous regarding any specific individual's guilt. The information in the vignette was limited mainly to what had happened and where, namely that a taxi driver had been shot in his car at a cul-de-sac. The vignette also mentioned that shoeprints had been found around the car, and that a pair of gloves was found near the crime scene. It mentioned

that the police had arrested a suspect who had a history of violent behaviour, but no evidence incriminating the suspect for this crime was included in the vignette.

In order to test the hypotheses, the vignette should not be biased towards guilt or innocence. Hence, no strongly incriminating or exculpatory information was included in the vignette. The vignette was pretested several times, and adapted according to the results each time. In the pretest of the final version of the case (N = 35), the average rating of the likelihood of the suspect being guilty was 47.34 (SD = 19.56) on a scale of 0 (*very unlikely*) to 100 (*very likely*).

4.2.3.2. Evidence. All participants were presented with four pieces of evidence after reading the case vignette. There were four types of evidence: eyewitness identification, hair comparison, shoeprint comparison, and CCTV footage. The evidence had been manipulated, such that each type of evidence existed in an incriminating version (e.g., the witness identified the suspect in a lineup) and an exculpatory version (e.g., the witness rejected the suspect in a lineup). The pieces of evidence were pretested extensively to ensure that they were perceived as approximately equally strong. In the final pretests (N = 78; with the exception of shoe prints which was added later, N = 35), the average strength ratings for the different pieces of evidence were all within 10 points from each other on a 100-point scale (for the detailed outcomes, see the Supplemental Materials; Appendix C).

4.2.3.3. Measures. Participants were asked to fill out a similar set of measures at three points in time: after reading the case vignette (Time 1), after having received the first set of evidence (Time 2), and after having received the second set of evidence (Time 3). They were asked to rate the extent to which they felt uncomfortable, uneasy, and bothered on scales ranging from 1 (*not at all*) to 7 (*very*; Elliott & Devine, 1994). The three ratings at each point in time were averaged to form composite measures of cognitive dissonance (Cronbach's α = .87, .90, and .91 for Time 1, Time 2, and Time 3, respectively). Participants were then asked to rate the likelihood that the suspect was guilty on a scale of 0 (*very unlikely*) to 100 (*very likely*). As court decisions on guilt in real life are dichotomous, participants were also asked whether they would convict the suspect (convict/acquit), and

how confident they were about their decision (0 = *not at all confident*, 100 = *very confident*). At the end of the experiment, participants were shown a list of the evidence they had received at Time 2 and Time 3, and were asked to rate how important each piece of evidence was for their decision regarding guilt (0 = *not at all important*, 100 = *very important*).

4.2.4. Procedure

Participants were provided with a link to the survey either through advertisements posters, university communication, or social media. The first page of the survey contained information about the study and participants were asked to provide informed consent. Those who consented were then directed to the demographics section of the questionnaire. Next, participants were shown the case vignette. After reading the case, they completed the measures for the first time (Time 1). Participants were then presented with the first set of two pieces of evidence. The valence of the presented pieces of evidence (i.e., both incriminating, both exonerating, or mixed) depended on the condition to which they had been randomly assigned. An overview of evidence combinations in each condition can be found in Table 1. Participants then completed the measures for the second time (Time 2). Next, they were presented with a second set of evidence, the valence of which was again dependent on participants' condition (see Table 1). Finally, they filled out the measures for the third time (Time 3), and then proceeded to rating the importance of the individual pieces of evidence. Participants were then debriefed through the presentation of written information about the purpose of the study, and were asked to leave their contact details in case they wanted a chance to win a 10 EUR voucher. Completion of the survey took on average 20 min 35 s (*SD* = 75 min 30 s).[1]

1 We suspect the large standard deviation can be explained by participants not closing the tab after they completed the survey.

4.3. Results

4.3.1. Preliminary Analyses

Participants could complete the questionnaire in either Dutch or English. Language did not interact with the dependent variables (analyses can be found in the Supplemental Material; Appendix C). It was therefore decided that language was not a confounding variable, and it was not included as a covariate in subsequent analyses. The data for the English and Dutch groups were combined.

4.3.2. Likelihood of Guilt

A 3 (time: Time 1 vs. Time 2 vs. Time 3) × 6 (condition: Mixed 1 vs. Mixed 2 vs. Ex/Inc vs. Inc/Ex vs. Ex/Ex vs. Inc/Inc) mixed ANOVA, with time as repeated measures, was conducted on participants' ratings of the likelihood of the suspect being guilty. The mean ratings for the different conditions at the different times can be found in Table 2.

Table 2

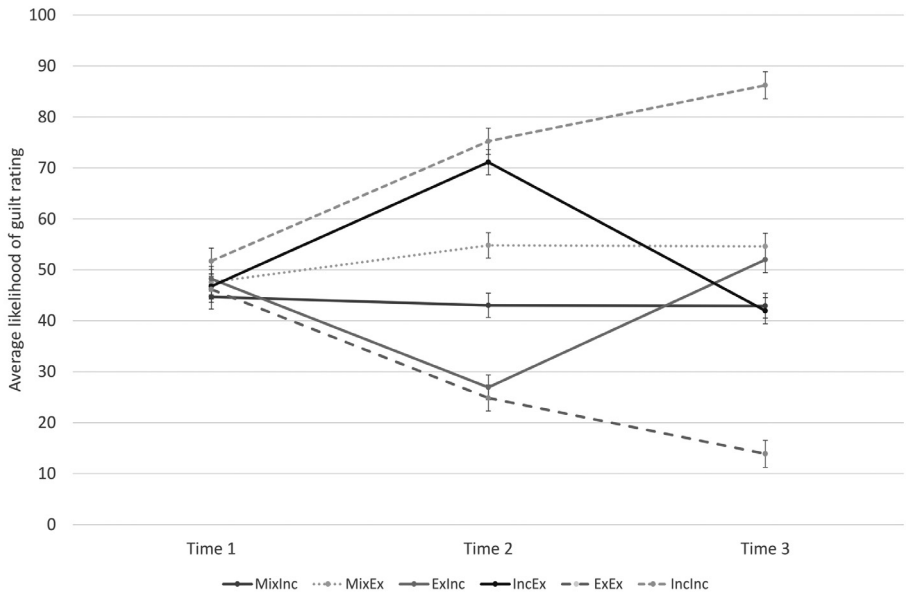
Means (and Standard Deviations) for Likelihood of Suspect Guilt at Different Times Across Conditions

Condition	Time 1	Time 2	Time 3
Mixed evidence			
Mixed 1	44.95 (21.32)	43.00 (22.59)	42.71 (24.12)
Mixed 2	47.63 (17.41)	54.79 (20.87)	54.55 (23.97)
Contradictory evidence			
Ex/Inc	47.97 (20.88)	26.96 (18.82)	52.50 (22.00)
Inc/Ex	46.38 (21.14)	71.16 (23.08)	42.26 (24.71)
Control conditions			
Ex/Ex	46.28 (21.63)	24.83 (18.48)	13.78 (16.03)
Inc/Inc	51.92 (19.69)	75.22 (16.59)	86.05 (12.10)

There was no significant main effect of time, $F(1.80, 719.64)^2 = 1.82$, $p = .167$, $\eta_p^2 = .005$, 90% CI [0.00, 0.02]. There was, however, a significant main effect of condition, $F(5, 401) = 43.32$, $p < .001$, $\eta_p^2 = .351$, 90% CI [0.28, 0.40] and a significant interaction effect between time and condition, $F(8.97, 719.64) = 73.40$, $p < .001$, $\eta_p^2 = .478$, 90% CI [0.43, 0.51] The nature of the interaction effect is illustrated in Figure 1. The interaction effect will be broken down into focused comparisons below.

Figure 1

Ratings for Likelihood of Suspect Guilt.



Note. Error bars represent standard error.

A linear trend analysis was conducted for the control groups (Ex/Ex and Inc/Inc), which had received uniformly incriminating or exonerating evidence in both the first and the second set of evidence. As would be expected, the

2 As the assumption of sphericity was not met, a Huyn-Feldt correction was applied, resulting in adjusted degrees of freedom.

linear trend was significant for both the Ex/Ex condition, $F(1, 63) = 118.72$, $p < .001$, $\eta_p^2 = .653$, 90% CI [0.53, 0.73] and the Inc/Inc condition, $F(1, 63) = 194.27$, $p < .001$, $\eta_p^2 = .755$, 90% CI [0.66, 0.81]. That is, participants made increasingly lower guilt rating in response to accumulating exonerating evidence and increasingly higher guilt ratings in response to accumulating incriminating evidence. This confirmed that the valence of the evidence was interpreted as intended.

Hypothesis 1 through 3 predicted that there would be an effect of the order in which participants viewed the evidence. Specifically, the likelihood of guilt at Time 3 for the condition which had first received incriminating evidence and then exonerating evidence (Inc/Ex) would be higher than for the conditions that had received mixed evidence at both times (H1). Conversely, the condition which first received exonerating evidence and then incriminating evidence (Ex/Inc) was expected to have a lower rating of likelihood of guilt at Time 3 than the mixed conditions (H2). We had also predicted that participants would show belief perseverance: Those who were first presented with exonerating evidence in the first set and incriminating evidence in the second set (Inc/Ex) were expected to have a lower rating of likelihood guilt at Time 3 than participants who received the evidence in the opposite order (Ex/Inc; H3).

A one-way ANOVA including the four experimental conditions was conducted on the likelihood of guilt at Time 3 to test these hypotheses.[3] There was a significant effect of condition, $F(3, 278) = 4.99$, $p = .002$, $\eta^2 = .051$, 90% CI [0.01, 0.09]. A planned contrast showed that the Inc/Ex group did not make significantly higher ratings of likelihood of guilt than the mixed groups combined, $t(275) = 1.83$, $p = .069$, $d = .26$, 95% CI [-0.02, 0.56], failing to support H1. A second planned contrast analysis showed that the Ex/Inc group did not make significantly lower ratings of likelihood of guilt than the mixed groups combined, $t(275) = -1.03$, $p = .305$, $d = -0.15$, 95% CI [0.00, 0.43], failing to support H2. A third planned contrast, comparing the two contradictory conditions,

3 This analysis differs from the preregistration. We realised that the registered analyses were overly complex to test this specific hypothesis, so we chose to do a between-groups ANOVA at Time 3 only, and using only the experimental conditions.

showed that participants in the Ex/Inc condition reported a significantly higher final likelihood of the suspect being guilty than participants in the Inc/Ex condition, $t(275) = -2.47$, $p = .014$, $d = -0.42$, 95% CI [0.11,0.75]. However, as we had predicted a difference in the opposite direction, H3 was not supported. Rather than belief perseverance, the latter finding is indicative of a recency effect (i.e., the evidence received last had the largest influence on the final guilt ratings).

4.3.3. Cognitive Dissonance

We also predicted that cognitive dissonance would differ between the conditions. Specifically, the contradictory conditions (Ex/Inc and Inc/Ex) were expected to report more cognitive dissonance at Time 3 than the mixed conditions (Mixed 1 and Mixed 2), which in turn were expected to report more cognitive dissonance than the control conditions (Ex/Ex and Inc/Inc). As the homogeneity of variance assumption was violated, a Kruskal-Wallis test was used.[4] The mean ratings for cognitive dissonance for the different conditions at all three times can be found in Table 3. A significant difference between the conditions was found, $H(5) = 24.15$, $p < .001$, $\varepsilon^2 = .081$, 90% CI [.070, .155][5]. A planned contrast analysis showed that the mixed conditions combined reported significantly more cognitive dissonance than the control conditions combined, $t(260.78) = -3.25$, $p = .001$, $d = .40$, 95% CI [0.17, 0.64] supporting H5. However, the mixed conditions combined did not differ significantly from the contradictory conditions combined, $t(268.42) = 1.18$, $p = .239$, $d = -.14$, 95% CI [-0.09, 0.37], failing to support H4.

4 This test also differed from the preregistration for the same reasons as described earlier. We realised a one-way ANOVA of cognitive dissonance at Time 3 only would be a more appropriate test of the hypothesis. Due to the violated assumption of homogeneity of variance, a Kruskal-Wallis test was ultimately used.

5 The confidence interval was computed using bootstrapping with 10,000 samples.

Table 3

Means (and Standard Deviations) for Cognitive Dissonance at Different Times across Conditions.

Condition	Time 1	Time 2	Time 3
Mixed Evidence			
Mixed 1	2.34 (1.38)	2.69 (1.62)	2.65 (1.53)
Mixed 2	2.00 (1.11)	2.05 (1.16)	2.13 (1.23)
Contradictory Evidence			
Ex/Inc	2.49 (1.50)	2.18 (1.27)	2.79 (1.61)
Inc/Ex	2.03 (1.11)	1.98 (1.00)	2.41 (1.40)
Control Conditions			
Ex/Ex	2.24 (1.13)	1.96 (1.16)	1.85 (1.20)
Inc/Inc	2.09 (1.32)	1.96 (1.22)	1.91 (1.19)

4.3.4. Exploratory Analysis

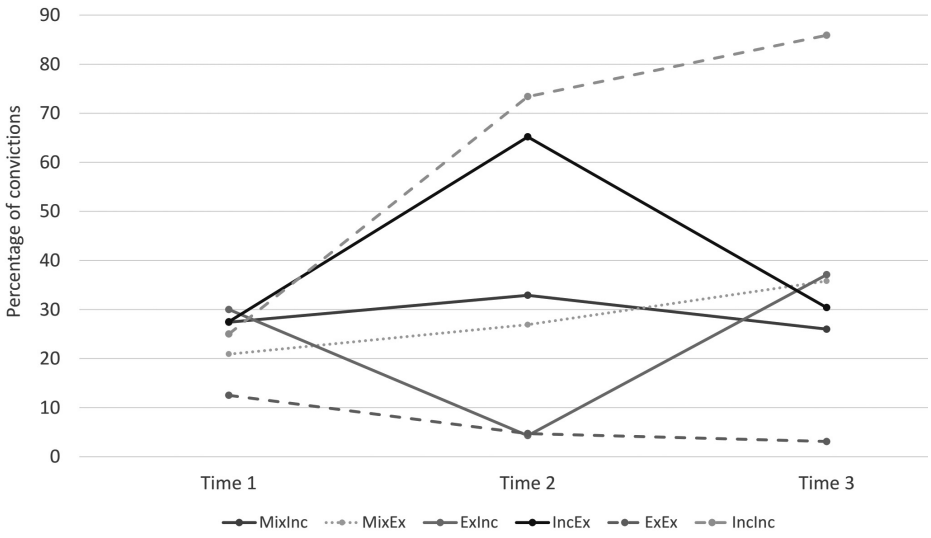
4.3.4.1. Dissonance and Belief Perseverance. We explored whether the cognitive dissonance experienced by participants at Time 3 was correlated with their change in likelihood of guilt rating from Time 2 to Time 3; that is, whether dissonance was a predictor of belief perseverance. The analysis is particularly relevant for the contradictory conditions, where belief perseverance would be indicated by a reluctance to change the ratings in response to the second set of (contradictory) evidence. Because the direction of change should logically differ between the conditions (upward in the Ex/Inc condition, downward in the Inc/Ex condition), they were analysed separately. The correlation was not significant in the Ex/Inc condition, $r = .049$, $p = .687$, or in the Inc/Ex condition, $r = -.014$, $p = .909$. The corresponding correlations in all conditions are reported in the Supplemental Materials (Appendix C).

4.3.4.2. Conviction Rates. Participants were asked whether they would convict the suspect. An overview of the conviction rates can be seen in Figure 2. To explore whether the decision to convict after seeing all the evidence differed because of the order in which the evidence was presented, a binary regression

analysis was conducted. As the control conditions did not receive the same evidence as the experimental conditions, they were not included in the analysis. The outcome variable was participants' decision on whether or not to convict the suspect at Time 3. Order of evidence presentation was not a significant predictor of the decision on whether to convict the suspect at Time 3, $\chi^2(3) = 2.58$, $p = .461$.

Figure 2

Percentage of Convictions in all Conditions at Different Times.



4.3.4.3. Importance Ratings. To explore whether the order in which the evidence was presented influenced the perceived importance of that evidence, the ratings for the pieces of evidence were analysed (see Supplemental Material for mean ratings for the evidence; Appendix C). As participants in the control conditions (conditions 5 and 6) only received either incriminating or exonerating evidence, they were not included in the analysis. A 2 x 4 mixed ANOVA was conducted, using the average importance ratings for both the incriminating and exonerating evidence as a within-subjects factor, and using experimental condition

as a between-subjects factor. There was a main effect of whether the evidence was incriminating or exonerating, $F(1, 275) = 11.33$, $p = .001$, $\eta_p^2 = .040$, 90% CI [0.01, 0.08]. The average rating of exonerating evidence ($M = 59.78$, $SD = 22.11$) was higher than the average importance rating for incriminating evidence ($M = 53.15$, $SD = 23.31$). However, there was no significant effect of condition, $F(3,275) = .41$, $p = .749$, $\eta_p^2 = .004$, 90% CI [0.00, 0.01], nor was there a significant interaction effect, $F(3,275) = 1.37$, $p = .252$, $\eta_p^2 = .015$, 90% CI [0.00, 0.04]. In sum, the exonerating (vs. incriminating) evidence was considered to be more important, and this did not differ depending on the order in which the evidence was presented.

4.4. Discussion

In the current study, we investigated whether the order in which incriminating and exonerating evidence is presented affects belief perseverance and cognitive dissonance. Contrary to our expectations, participants did not show belief perseverance in their ratings of the likelihood of the suspect being guilty. Instead, the effect of order on ratings for likelihood of guilt appeared to be a recency effect. There was also no difference between the mixed evidence conditions and the contradictory evidence conditions for cognitive dissonance, again failing to support our hypotheses.

Participants in all conditions seemed to appropriately adjust their ratings of guilt likelihood in response to the evidence they were presented with. For the mixed conditions, likelihood of guilt stayed approximately neutral. For the contradicting conditions, participants' rating of likelihood of guilt increased when they were presented with incriminating evidence and decreased when they were presented with exonerating evidence. In contrast, other researchers have found that participants who expressed a belief in the suspect's guilt subsequently focused more disproportionally on the suspect than those who did not express a belief in guilt (Rassin et al., 2010; O'Brien, 2009). Our findings, however, seem to be in line with the findings by Price and Dahl (2014) and Charman et al. (2016), as the last piece of evidence received by the participants seemed to disproportionally affect

their final rating of likelihood of guilt; those who saw incriminating evidence last had a higher final rating of likelihood of guilt than those who saw exonerating evidence last. Participants seemed to be unaware of the recency effect, as they did not rate the final pieces of evidence to be more important than previous pieces. Moreover, the evaluation of evidence did not seem to be affected by the order in which it was presented, as there was no difference in the importance ratings of the evidence between the different conditions.

The cognitive dissonance ratings also did not seem to be affected by the order in which the evidence was presented. Furthermore, we found no significant correlation between the cognitive dissonance reported by participants and the change in their rating of likelihood of guilt. Favouring information that supports one's belief is a strategy to avoid or reduce cognitive dissonance (Festinger, 1957; Jonas et al., 2008). However, if participants did not experience cognitive dissonance in response to the contradicting information, they would also have no need to adhere to a specific belief in order to reduce cognitive dissonance. The generally low reported levels of cognitive dissonance are therefore in agreement with the lack of belief perseverance that we observed.

4.4.1. Recency Effect

Contrary to the theory of belief perseverance, but in line with previous research (Charman et al., 2016; Costabile & Klein, 2005), we observed a recency effect in the ratings of likelihood of guilt. The observed recency effect did not, however, appear to affect participants' decision to convict or acquit the suspect: there was no significant difference between the conditions for final conviction rates. As conviction rates were the main measure used by Costabile and Klein (2005), it is unclear why the recency effect in our study did not affect the conviction rates. Costabile and Klein (2005) suggested that the overall evaluation of information is delayed until a time when one is specifically asked for a decision, at which point the decision is made based on the evidence that is most readily accessible (i.e., the evidence received last). In our study, participants

were asked for an overall evaluation of the information, namely whether they would convict the suspect, at three different times. In doing so, they integrated the evidence they had seen until that point. The integration might therefore have taken place in several steps rather than all at once, which could have limited the influence of the evidence that was most easily accessible. It could be argued that, despite the last evidence disproportionally affecting the likelihood of guilt rating, the effect was not strong enough to produce a change in participants' decision to convict or acquit the defendant. Such an interpretation would suggest that the practical consequences of a recency effect may be limited.

Another explanation can be found in the study by Kerstholt and Jackson (1998). They compared the integration of evidence when participants were asked to give a probability rating of guilt after each witness statement as opposed to after reading all the evidence. They found a recency effect in the former condition. According to their explanation, the recency effect is due to an anchoring-and-adjustment process, whereby new information is evaluated in relation to the general impression of previous evidence, and more weight is attached to the last evidence in adjusting the anchor. Relating their explanation to the findings in the current study, it could be that participants in the contradictory-evidence conditions adjusted their rating towards the impression of the evidence presented at Time 3 and away from the opposite impression created at Time 2. Due to the considerable difference between the earlier impression and the latter impression, however, the latter evidence may not have created a sufficiently strong impression of guilt or innocence to affect conviction rates.

Both of the explanations offered above relate to the difference between the rating of the likelihood of the suspect being guilty and the conviction rates. The discrepancy between our findings and the existing literature (e.g. Charman et al., 2016) suggest the relation between perceived likelihood of guilt and the absolute decision on guilt may be more complex than previously thought. If researched further, additional data on this relationship could add to the general literature on legal decision-making.

4.4.2. Evidence Evaluation

Based on the changes in rated likelihood of guilt in response to the presentation of evidence, we can conclude that participants were not disproportionally influenced by a prior belief in their evaluation of the evidence. There was also no effect of condition on the importance ratings of the individual pieces of evidence. That finding is not in line with the finding by Charman et al. (2016), who found that the initial piece of evidence viewed by participants affected their evaluation of the subsequent evidence. It is also not in line with the findings by Price and Dahl (2014), who found the recency effect extended to the evaluation of the individual pieces of evidence.

One unexpected finding was that participants rated the exonerating evidence as more important than the incriminating evidence. Marksteiner and colleagues (2011) found that only police trainees with a prior belief in the suspect's innocence rated incriminating and exonerating evidence as equally reliable, whereas those with a guilty hypothesis rated incriminating evidence as more reliable than exonerating evidence. Furthermore, the pretests of the evidence used in the current study also showed a slight guilt bias; the incriminating evidence was consistently rated as somewhat stronger than its otherwise identical exonerating equivalent. We therefore expected the incriminating evidence to be more important according to participants. It should be noted that the evaluation of the evidence in earlier research has often been done by asking participants to rate the reliability or strength of the evidence itself, whereas we asked them to rate the importance of the evidence for their decision whether to convict or acquit the defendant. As participants were asked about the importance of the evidence retroactively, the question related more to how they integrated the evidence to reach their decision. That may have limited the effect of either the last piece or first piece of evidence on the evaluation of the individual evidence. By examining the importance of the evidence that contributed to the participant's decision, the influence of a single piece of evidence may have been prevented. The contribution of the current study to the existing literature suggests that potential effect warrants further research.

The fact that participants in the main study rated exonerating evidence as more important might also indicate an innocence bias in the participants who took part. That could be due to the sampled population. When Ask and Granhag (2005) found that professional investigators were less receptive to a suggested alternative scenario than lay participants, one explanation for their finding was that professional investigators are already likely to consider alternative scenarios by default. Similarly to police officers, law students may also spontaneously consider different scenarios, or may even have a bias towards innocence. After all, law students are not only the judges of the future, but also the defence lawyers of the future. That could explain the higher importance ratings for the exonerating evidence, as well as provide an explanation as to why, despite the higher rating of likelihood guilt, participants who saw incriminating evidence last were no more likely to convict the suspect. In order to determine whether the proposed explanation affected the findings, the current study could be conducted with practicing lawyers.

4.4.3. Implications and Future Research

The fact that we did not observe belief perseverance, and the fact that the recency effect did not affect conviction rates leads to a tentative optimistic conclusion based on the data; the order in which evidence is presented may not strongly influence the occurrence of biased decision-making. However, further research in a more realistic context is necessary. Furthermore, it must also be remembered that treatment and consideration of cases is far more elaborate in real trials than in the current study, which likely also affects the evaluation of evidence. In addition, real life case files are also likely to contain mainly incriminating information when a case is brought to trial (Crombag, 2017). As several researchers have previously found that the presentation of guilt-biased information influences the evaluation of later evidence (e.g., Ask et al., 2011, Ask et al., 2008, Greenspan & Scurich, 2016), a replication of the current study using an incriminating case file may also provide further insights.

The mixed conditions in the current study differed from the majority of previous research. The mixed condition which started with the incriminating evidence can be argued to be more similar to trial proceedings in the Netherlands. Van der Post and Van Toor (2019) summarised the trial proceedings as following an incriminating–exonerating–incriminating–exonerating sequence. The trial starts with the prosecutor presenting the indictment and the questioning of the suspect about the evidence, followed by the defence lawyer's plea. The prosecutor then responds to the lawyer, after which the lawyer responds again and the suspect gets to speak last (Van der Post & Van Toor, 2019). According to the results of the current study, that order should not affect the perception of the evidence, as the mixed conditions did not differ from the contradicting conditions in their rating of likelihood of guilt. However, the complexity of a trial can hardly be compared to the procedure of the current study. It would therefore be beneficial to attempt to replicate the current findings in a trial setting with higher ecological validity.

There are also some differences between the procedure used and real-life proceedings which limits the application of our findings. For instance, Kerstholt and Jackson (1998) found that participants who were asked to judge the defendant's guilt after seeing all the evidence showed a recency effect when background information was provided, but a primacy effect when no background information was provided. In our study, participants were not provided with background information about the suspect. In reality, judges in the Netherlands receive the case file and then have about a week to prepare for the trial (Van der Post & Van Toor, 2019). Such background information would, according to Kerstholt and Jackson (1998), contribute to a recency effect. Furthermore, the time delay between the presentation of evidence and having to make a decision can also have an impact on possible order effects. More time between the different pieces of information that are being presented has been associated with an increased recency effect, whereas more time between the final presentation and the recall has been associated with a decreased recency effect (Insko, 1964; Stout et al., 2005). In the Netherlands, smaller criminal cases processed by a single judge will be decided immediately

at the end of the trial. More serious or complicated cases with multiple judges will usually be discussed by the judges immediately after the trial or at the end of the day (Van der Post & Van Toor, 2019). A decision is then written, and the judgement is usually announced two weeks after the trial. Therefore, the current study was most similar to the trial by the single judge, as participants were asked to make a decision on their own immediately after seeing the evidence. Replicating the study with a longer time delay, and including background information, could therefore provide insight into possible order effects in a greater variation of proceedings.

There are a few limitations to the current study. Perhaps the biggest limitation is the use of law students instead of actual judges. Cognitive dissonance may be particularly likely to arise when the belief in question is considered important. It can be expected that judges are more invested in cases they work on than law students are in an experimental study. Furthermore, the training and experience of judges may cause them to respond differently than law students. That would be in line with the findings by Schmittat and Englich (2016), who found that criminal law experts showed less preference for confirming information in a criminal law case than did experts in other areas of law and laypeople. In addition to the need to determine whether law students and defence lawyers have a preference for alternative scenarios, the question of whether and how the consideration of clearly exonerating evidence differs between the different parties at trial also warrants investigation. Furthermore, based on the current study, it also seems that findings on order effects among mock jurors (e.g., Costabile & Klein, 2005) may not generalise to populations which have received legal training.

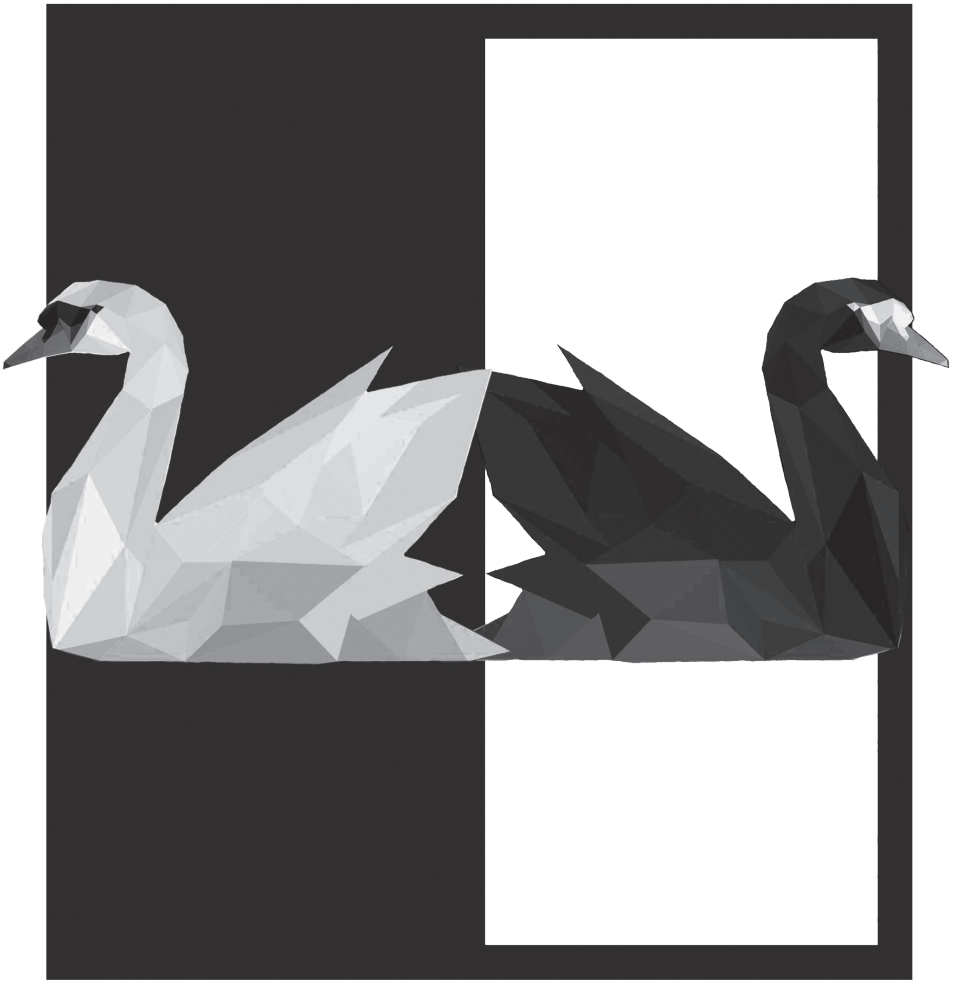
A second possible limitation is the fact that the study was conducted online. While this sped up the data collection process and extended the achievable sample size, we cannot be sure how attentive participants were while taking the survey. We did, however, exclude participants who took the questionnaire very quickly in an attempt to filter out participants who rushed through the survey or who could not have read the material carefully. A third limitation of the study is that only one type of crime was used, namely a murder case. The case we used

here has been used in multiple other studies (Ask et al., 2011; Ask et al., 2008; Marksteiner et al., 2011), although it was adapted slightly for the purpose of the current study. Other studies on order effects (e.g., Charman et al., 2016) have also used murder cases. It therefore seems unlikely that the choice of case can account for the unexpected results in the current study. Nevertheless, it may be beneficial for future research to include several types of crimes.

A final limitation that should be considered is related to the measurement of cognitive dissonance. We expected that the simultaneous presentation of mixed evidence would not give rise to substantial cognitive dissonance. However, the mixed conditions experienced stronger dissonance than anticipated, resulting in dissonance not too different from that caused by the contradicting evidence in the other conditions. There is also an inherent difficulty to measuring dissonance. As it is an internal feeling, there is a risk that simply measuring dissonance could interfere with participants' experience of dissonance. Nevertheless, the dissonance measure has been used successfully in previous related research (Ask et al., 2011). Elliot and Devine (1994) created the measure based on Festinger's (1957) description of the state of dissonance and on conceptually related research on affective responses to manipulations. Therefore, we considered this particular instrument to be the most valid available measure of cognitive dissonance.

4.4.4. Conclusion

While the hypotheses of the current study were not supported, we observed a recency effect in the ratings of likelihood of guilt. The recency effect did not extend to the decision whether or not to convict the suspect, and also did not influence the perceived importance of the evidence. Thus, despite the final rating of likelihood of guilt shifting towards the last piece of evidence that participants saw, participants were not sufficiently influenced by the evidence to cause a difference in conviction rates. The recency effect therefore did not seem to undermine the impact of the evidence presented earlier. It would be meaningful to attempt to replicate the current findings using a more ecologically valid procedure and sample.



CHAPTER 5

COGNITIVE BIAS AND ACCOUNTABILITY IN LEGAL DECISION-MAKING

An adapted version of this chapter has been submitted for publication as:
Maegherman, E., Ask, K., Horselenberg, R., & Van Koppen, P. J. (2020) *Cognitive bias and accountability in legal decision-making* [Manuscript submitted for publication] Department of Criminal Law, Maastricht University.

Abstract

Having to explain a decision has often been found to have a positive effect on the quality of a decision, although the effect varies depending on a number of factors. In the current study, we aimed to determine whether the different accountability requirements for judges in the Netherlands (i.e., justification) and Germany (i.e., explication) affect evidence use and integration. Those requirements were compared to instructions based on the principle of falsification and a control condition. Participants (N = 173) decided on the defendant's guilt in a murder case vignette and explained their decision according to the instructions. The explication and falsification (but not the justification) instructions increased the use of exonerating evidence. There was no significant difference between the groups in guilt perception. The use of exonerating evidence was a significant positive predictor of acquittal rates. The findings have positive, although complex, implications for the use of accountability instructions in practice.

5.1. Introduction

Although the process of legal decision-making has been the subject of a variety of theoretical explanations as well as experimental research, the insight into how judges reach a final decision remains limited. Some aspects of the decision-making process are known, as they are prescribed by law. A requirement for judges in several countries is that they are, to a certain extent, required to account for their decision (e.g., Art. 359 and 360 Dutch Code of Criminal Procedure [DCCP]; Mevis, 2019). In previous research on decision accountability, researchers have suggested that such a requirement could substantially alter the decision-making process (Lerner & Tetlock, 2003). In the current study, we investigated whether variations in the instruction on how to account for a decision affect the evidence considered and the decision made on the guilt of a defendant.

5.1.1. Reasoned Judicial Decisions

As a judge can almost never know for sure what exactly happened, an inherent leap is required in order for them to become sufficiently convinced about what happened based on the information provided in the evidence. One of the elusive aspects of legal decision-making is how that leap is made. The need and requirements for explaining a decision differ between the various legal systems, but one common expectation is that the explanation will provide some sort of insight into the judicial decision on guilt in criminal legal proceedings. The most important question to be answered by the judge is whether the suspect committed the crime he is accused of (Dreissen, 2007). In order to answer this question, judges in the Netherlands will first study the case file, which is likely to consist of mainly incriminating information, and will then be presented with the prosecution's and the defence's arguments at trial (Crombag, 2017; Verbaan, 2016). The reasoned decision should make it clear that the rules regulating the use of evidence were followed. It can also be seen as an explanation of why the

judge was convinced beyond a reasonable doubt that the accused committed the crime (Dreissen, 2007).

Besides the requirements that the reasoned decision has to fulfil, there are several additional functions for why judges in the Netherlands must explain their decisions. Firstly, the explanation of their decision acts as justification for the punishment that follows for the convicted individual. Secondly, the reasoned decision is used to account for the decision to the general public. Thirdly, it informs the various parties involved in the legal proceedings. Lastly, the reasoned decision can serve as a potential quality control by other legal instances, such as the Supreme Court, although that rarely happens in practice (Verbaan, 2016; Dreissen, 2007). Furthermore, it has been argued by Gommer (2007) that there is a need for requiring an explanation due to the potential influence of thought-processes the decision-maker may not be aware of, such as biases. In theory, the explanation serves as a "rational reconstruction" of what was considered by the judge for the decision (Gommer, 2007).

5.1.2. National Differences in Accountability Requirements

Different legal systems incorporate different instructions on how a decision should be accounted for. Scholars have compared the content of the Dutch requirement to the German requirement for explaining a decision (Dreissen, 2007; Mevis, 2019). Although there is little difference between the codes of criminal procedure in the Netherlands and Germany on that issue, the literature on the explanation requirements makes it clear that the German system imposes stricter requirements on the judge (Mevis, 2019; Simmelink, 2001; Dreissen, 2007). Whereas the German instructions could be interpreted as requiring an *explication*, the Dutch instructions could be interpreted as requiring a *justification* of the decision. In the German system, the judge has to account for their selection and evaluation of evidence, and to pay attention to facts that indicate an alternative, but not accepted, version of events (Mevis, 2019; Dreissen, 2007). Furthermore, there are specific requirements of evidence evaluation. For

instance, in case of contradicting witness statements, the judge has to consider how both statements came about, as well as to explain the discrepancies between them. In the written decision, the judge will have to account for the grounds of his reliability judgement. Overall, the German judge is required to provide a more in-depth explanation of the decision than the Dutch judge. In doing so, the judge shows the decision was made by a professional with integrity rather than by a purely subjective individual (Mevis, 2019).

In the Netherlands, the requirements imposed on the judge to explain or motivate his decision are limited, due to the integrity and professionalism inherently expected of a judge (Mevis, 2019). The explanation provided by the judge does not have to be a reflection of the discussion or consideration that led to the decision. It suffices if the explanation contains arguments which, taken together, *justify* the decision that was rendered (Reijntjes & Reijntjes-Wendenburg, 2018). The point of view that the selection and evaluation of evidence does not require motivation, with a few exceptions, is in stark contrast to the extensive requirements in the German system (Mevis, 2019). Although article 360 of the Dutch Code of Criminal Procedure requires that the judge explicitly motivates why he considers certain evidence to be reliable, it is limited to evidence where the reliability is questionable (e.g., vulnerable or anonymous witnesses). Compared to the Dutch standards, the German judge has an extensive duty to motivate the decision—the written decision not only needs to include the proven fact and the evidence used, but also needs to explain the selection and evaluation of evidence (Simmelink, 2001). The Supreme Court of the Netherlands appears to be lenient in enforcing the rules regarding the reasoned decision provided by the judge (Dreissen, 2007). The review of the decision by the Supreme Court remains limited following a change in the Dutch Code of Criminal Procedure in 2005; the judge now explicitly has to explain why their decision differs from the substantiated arguments raised by either the prosecution or the defence. Thereby, the extent of the reasoned decision becomes dependent on the points raised by one of the parties (Dreissen, 2007). The differences between the Netherlands and Germany

in their requirements for the reasoned decision raise the question of how these differences affect judges' reasoning with evidence.

5.1.3. Impact of Accountability on Reasoned Decisions

The need to account for the decision on guilt or innocence of the suspect thus appears to differ between legal systems. In trying to determine the effects of such differences, it is important to consider existing research on the effects of the requirement to explain one's decision. Researchers have identified several ways in which such accountability can affect the decision-making process (Lerner & Tetlock, 2003). A key aspect of accountability, which determines its effectiveness in reducing cognitive bias, is whether the requirement to account for a decision was known prior to making the decision. Prior accountability is thought to encourage exploratory reasoning and making an optimal judgement, whereas post-decisional accountability has been found to increase confirmatory and self-justifying reasoning (Lerner & Tetlock, 1999). For judges, it is very clearly established prior to reading a case that an explanation for their decision will need to be provided. Taking that into consideration, it seems that the instructions to motivate a decision prior to making a decision can encourage unbiased thinking, and facilitate a better judgement by the decision-maker.

One of the frequently considered factors of accountability is the distinction between having to explain the decision versus having to explain the decision-making process (Tetlock, 1985). These are known as *outcome accountability* and *process accountability*, respectively. Although process accountability is often found to be associated with better decision-making, some researchers have suggested it may depend on the type of task, with the beneficial effect being restricted to elemental tasks involving linear relations between the cues and the outcome (De Langhe et al., 2011; Hall et al., 2015; Lerner & Tetlock, 2003). If we consider those findings in light of the explanation required of judges in the Netherlands, it appears that their accountability is focused more on explaining the decision itself than on explaining the decision-making process that lead to that decision.

In fact, the Dutch Supreme Court has ruled that the reasoned decision does not have to reflect the evidence that was considered, but merely the evidence that the final decision could reasonably be based on. The reasoned decision is therefore not a valid reflection of the decision-making process but rather is focused on outcome accountability (Reijntjes & Reijntjes-Wendenburg, 2018).

Another factor which has been found to moderate the effects of accountability on the decision-making process is the audience to whom the decision needs to be accounted for. Researchers have found evidence that the accountable persons shift their opinion towards the perceived opinion of the audience. For instance, students who were judging a cheating case gave harsher punishments to the perpetrator when they thought they had to explain their decision to the professor who reported the cheating than when they thought they had to explain it to the accused student (Pennington & Schlenker, 1999). However, research on multiple audiences is lacking (Hall et al., 2015). In the case of judges, the audience may hold a range of opinions. For instance, the decision will likely be read by the defendant and their relatives, but possibly also by the complainant and their relatives, as well as the public and other judges. Furthermore, the court of appeal may also read it. In addition to the audience, the source of information may also have an influence, as bias is thought to be amplified when the information is provided to the decision maker by someone who knows what is relevant to the decision at hand (Lerner & Tetlock, 1999).

Researchers investigating accountability have mainly focused on other areas of decision-making, and little research has been conducted into accountability in the context of legal decision-making. Tetlock (1983) investigated whether the influence of an initial impression of guilt can be affected by prior accountability. He found that those who initially received evidence against the defendant first were more likely to find him guilty, but that this primacy effect was reduced by prior accountability. Therefore, prior accountability seems to be able to prevent an initial belief from biasing a decision on guilt, which has obvious positive implications for the requirement of judges to explain their decision.

5.1.4. Assessing Quality in Legal Decision-Making

Scholars have suggested that forcing judges to substantiate their decisions can enhance the accuracy of legal decision-making, by ensuring that the decision is not based on irrelevant information or speculation (Cohen, 2015). This notion, however, does not accommodate the intricate effects of accountability on decision-making as demonstrated by the psychological research reviewed above. The lack of understanding concerning the effects of accountability on legal decision-making may be due to the difficulty in assessing what constitutes a good decision in the legal context. Some elements of what the accountability literature considers to be important for a good decision can also be seen in the context of decision-making by judges, such as the notion of impartiality. However, in actual legal decision-making, an objective ground truth is often not available, which makes it difficult to determine the quality of the decision.

A reasoned decision can give some insight into what was considered in making the decision. For instance, indicators of confirmation bias in a reasoned decision, such as an excessive focus on incriminating evidence while ignoring exonerating evidence, would suggest that the decision-making process may have deviated from its goal of determining the truth. As the determination of the truth is generally considered the aim of criminal proceedings (Cleiren, 2008; De Keijser, 2017; Crombag, 2017), written decisions containing indicators of confirmation bias could be considered as an arguably worse decision, as it would suggest a focus on an existing belief rather than finding the truth.

Based on previous research, the active consideration of alternative scenarios can mitigate the influence of a prior belief (O'Brien, 2009; Rassin, 2018). Therefore, explanations that consider alternative scenarios can be considered indicative of a less biased process of decision-making. Trying to disprove an existing idea, known as falsification, can also be considered an important process when trying to determine what most likely happened. Several serious failed attempts at disproving a theory make it more likely that the theory is true (Crombag, Van Koppen, & Wagenaar, 2006). Falsification is closely related to the consideration

of alternative scenarios, as evidence that disproves one scenario may confirm another scenario. Furthermore, trying to find a good alternative scenario for the available evidence can also be considered part of attempting falsification (Van Koppen & Mackor, 2019). Although consideration of exonerating evidence and alternative scenarios remain indirect measures of the decision-making process, they can provide insight into whether the evidence considered for the decision, and thus the decision-making process, differs depending on the instruction given to account for the decision.

5.1.5. The Current Study

In the current study, we aimed to investigate whether prior instructions to account for a decision affect the legal decision-making process. In order to do so, participants were provided with one of four instructions before reading a vignette of a murder case and were then asked to make a reasoned decision on the guilt of the defendant. In the *justification* condition (based on the Dutch Code of Criminal Procedure), participants were asked to mention evidence that supported their decision, while in the *explication* condition (based on the German Code of Criminal Procedure) participants were asked to show that they had considered evidence both for and against their decision. In the *falsification* condition, participants were asked to describe the different possible versions of events and how they decided on the most plausible version by excluding the alternatives. In the fourth condition, which was considered the *control* condition, participants only received the general instruction to explain their decision. After reporting and explaining their decision, participants were asked to rate the individual pieces of evidence in the case for how incriminating or exonerating they found them to be.

We formulated and preregistered the following hypotheses:

H1: Those in the justification condition were expected to use less exonerating evidence in their justification of the decision than those in the explication or the falsification condition, but more than those in the control condition.

H2: Those in the justification condition were expected to consider fewer scenarios than those in the explication or falsification condition, while the control group was expected to consider fewer scenarios than the three experimental conditions.

H3: (a) The justification condition was expected to have a higher conviction rate than the explication or falsification group, (b) but the control condition was expected to have a higher conviction rate than the three experimental conditions. (c) The average rating of guilt was also expected to be higher in the justification condition than in the explication and falsification conditions, (d) while the control condition was expected to have a higher rating of guilt than the three experimental conditions.

H4: The amount of exonerating evidence mentioned in the written decision was expected to be a significant negative predictor of the conviction rate of the suspect.

The preregistration for the study can be found at: https://osf.io/fc962/?view_only=5746019c60bf4a4e84ef103627a0a0e8

5.2. Methods

5.2.1 Participants

Participants were recruited online using Qualtrics and using Amazon MTurk, as well as via advertisements through social media. A power analysis conducted in G*Power (v3.1; Faul et al., 2007), using a medium effect size (f = .25) and power of 0.8 resulted in a required sample size of 179 participants. To allow for potential exclusions, considerably more responses were collected. A large number of participants (n = 366) did not answer the control questions about the instructions correctly and were screened out from the survey at an early stage. Responses were also excluded for incorrect answers to the control questions about the case (n = 33) or to the attention checks (n = 37), and for open-ended answers which we suspected were not genuine (e.g., bots, duplicate responses; n = 49). Incorrect answers to the control questions and the attention checks

had been preregistered as exclusion criteria. One of the control questions was ultimately not used as an exclusion criterion because the answer was not clear enough from the vignette. Another eight participants were excluded because their rating for likelihood of guilt differed from the sample median by more than three times the absolute median deviation (Leys, Ley, Klein, Bernard, & Licata, 2013), which had also been pre-registered as an exclusion criterion. The final sample consisted of 173 participants. Participants received compensation through MTurk. Participants' mean age was 31 years old (*SD* = 11). The majority of participants (58%) was female.

5.2.2. Materials

5.2.2.1. Instructions. Participants were randomly assigned to one of four conditions, each of which received a different instruction on how to motivate the decision they had made (Appendix D). These were the *justification* condition (based on the Dutch Code of Criminal Procedure), the *explication* condition (based on the German code of criminal procedure), the *falsification* condition (based on the principle of falsification), and the *control* condition (in which participants were given only minimal instructions). The various instructions were constructed after consultation of the literature on the requirements for judges in the different countries to account for their decision. They had also been pretested to ensure that they were as comprehensible as possible.

5.2.2.2. Practice Vignette. In order to familiarise participants with the instructions, they received a practice vignette and were asked to make a decision and motivate it. The case concerned a burglary, where a suspect had been charged for the crime, but was accusing someone else. The example was a simple task which allowed participants to practice applying the instructions.

5.2.2.3. Case Vignette. Participants were then presented with one vignette of a fictional murder case. In the case, Emma Miller claimed to have found her husband James dead when she arrived home from seeing her friend. Emma was covered in blood when the police arrived. The case contained information

about James having had an extra-marital affair, and that Emma knew about the affair. Emma was described as the main suspect based on the evidence against her. However, the case also contained a few indications of an alternative scenario. The pieces of evidence in the case were pretested for the extent to which they were perceived as incriminating or exonerating. As intended, the case was perceived as indicating that Emma was guilty of killing James, with an average likelihood of guilt rating of 69.1 (SD = 16.9) on a 0–100 scale in the pretest (N = 71).

5.2.3. Measures

5.2.3.1. Case Judgments. After writing their reasoned decision, participants were asked to rate how likely it is that the main suspect, Emma, killed James, using a visual analogue scale from 0 (*not at all likely*) to 100 (*very likely*). Following that rating, participants were asked whether or not they would convict Emma for murdering James by selecting one of two options (*acquit/convict*). After making their decision, participants then had to rate how confident they were about their conviction decision on a visual analogue scale from 0 (*not at all confident*) to 100 (*very confident*).

5.2.3.2. Valence Ratings of Evidence. Participants were asked how exonerating or incriminating they found individual pieces of evidence to be. They did so by using a visual analogue scale from 0 to 100, where 0 means completely exonerating and 100 means completely incriminating. In order not to influence participants' judgments in either direction, the starting position of the slider was set to 50 when participants were first presented with the scale. These ratings were not included in the hypotheses or used in the main analyses, but can be found in the Supplemental Materials (Appendix E). They were not included in the main analyses as it could not be determined whether participants used the evidence as predicted by the pretest results.

5.2.4. Procedure

All participants completed the study online using Qualtrics. Participants were first welcomed to the study and provided with some information about the study, such as that they would have to judge the guilt of the defendant and have to explain their decision. They then provided informed consent before starting the study. Participants first filled in a short demographic section, including their age and educational background. In the next section, participants were randomized to one of the four experimental conditions and were given the instructions to explain their decision according to the condition. Multiple choice control questions about what participants were asked to do according to the instructions were included to ensure that the instructions were correctly understood by participants. If participants did not answer all control questions correctly, they were directed back to the instructions and could then attempt the questions again. There was one control question for each element of the instruction, which resulted in two questions for the control condition, three questions for the justification and falsification condition, and four questions for the explication condition. Participants could attempt the questions twice in the control condition, three times in the justification and falsification condition, and four times in the explication condition. If, after the final attempt, they still did not answer all questions correctly, they were taken to the end of the survey and did not continue to the actual study.

The participants were then given a short practice vignette depicting a burglary case and asked to decide on the guilt of the defendant and explain their decision, thereby familiarizing themselves with the instructions. Participants were told that they were required to write a reasoned decision after reading the actual case vignette. In order to increase their sense of accountability, participants were told that their explanation would be reviewed by a panel of judges to determine how well they explained their decision according to the instructions. In the next section, participants were presented with the actual case vignette, and were asked to write a reasoned decision about the

case. While writing their decision, participants were able to refer back to the case description which was presented on the same page. In the final section, participants filled in the measures described above. Here, again, they could revisit the case description. Participants were thanked for taking part in the study and received further information about the aims of the study. The median response time was 35 minutes and 8 seconds.

5.2.5. Coding of the Reasoned Decisions

The coding of the evidence in the reasoned decisions was based on how the evidence was used by participants. The categories established by the pretest (see Supplemental Material; Appendix E) were not used for the final analyses, as the way in which the evidence was actually used was considered to be more representative of participants' reasoning process than how the evidence has previously been defined. Using the evidence according to the coding categories allowed for better incorporation of participants' interpretation of the evidence into the conclusions drawn based on the data. The coding was ultimately divided into three categories: incriminating evidence, neutral evidence, and exonerating evidence.

All the responses were coded by one coder. A second coder coded 21 of the responses (12%) in order to assess the interrater reliability of the coding. The intra-class correlation coefficient (ICC) for a two-way, random, single-measures, consistency analysis was conducted for the categories of exonerating, incriminating and neutral evidence, as used by the participants. The ICCs for the categories of evidence are reported in Table 1. As only the coded exonerating evidence was used to test the study's hypotheses, the ICCs were considered acceptable.

Finally, the scenarios mentioned by the participants were coded according to the implicated perpetrator: Emma, James' mistress, or "other". An intra-class correlation coefficient for a two-way, random, single-measures (consistency) analysis was conducted for each of the scenario categories (see Table 1). The ICCs for the scenario categories were good.

Table 1

ICC for the Coded Evidence and Scenarios in the Written Decisions

Category	ICC	95% CI lower bound	95% CI upper bound
Incriminating evidence	.949	.906	.972
Neutral evidence	.581	.340	.751
Exonerating evidence	.911	.841	.951
Scenarios implicating Emma	.816	.682	.896
Scenarios implicating James' mistress	.859	.752	.921
Scenarios implicating other perpetrators	.792	.645	.882

5.2.6. Design and Analysis

The experiment included four independent conditions to which participants were randomly allocated. A number of dependent variables were used to test the hypotheses. A one-way between-groups ANOVA, using the amount of exonerating evidence used in the reasoned decisions as dependent variable, had been planned to test H1. A similar one-way ANOVA, using the number of scenarios mentioned by participants as dependent variable, had been planned to test H2. A Pearson's chi-square analysis had been planned to determine whether the conditions differed in their decision to convict the participant, as predicted in H3. Furthermore, a one-way between-groups ANOVA, using the rating of likelihood of guilt as dependent variable, had been planned to test H3. Lastly, a point-biserial correlation coefficient had been planned to determine whether the amount of evidence coded as exonerating in the written decisions was significantly associated with participants' decision to convict or acquit the main suspect (H4).

5.3. Results

5.3.1. Use of Exonerating Evidence (H1)

A one-way between-groups ANOVA was conducted to compare the amount of evidence coded as exonerating for each of the conditions (for means in each condition, see Table 2). As the assumption of homogeneity of variance was

violated, the analysis differed from the pre-registration, and a Welch ANOVA was conducted instead. A significant difference between the conditions was found, $F(3, 92.26) = 4.64$, $p = .005$, $\eta_p^2 = .131$, 90% CI [.026, .219]. A planned contrast was used to compare the mention of exonerating evidence in the justification condition to the combined falsification and explication conditions. There was no significant difference, $t(89.6) = 1.52$, $p = .132$, Hedges' $g = -0.28$, 95% CI [-0.64, 0.09]. Another planned contrast was conducted to compare the exonerating evidence mentioned in the justification condition to the exonerating evidence mentioned in the control condition. No significant difference was found, $t(78.1) = -1.55$, $p = .125$, Hedges' $g = 0.33$, 95% CI [-0.09, 0.75]. Hypothesis 1 was therefore not supported.

An additional exploratory contrast was conducted to test whether there was a significant difference between the combined explication and falsification conditions compared to the control condition for the mention of exonerating evidence. A significant difference was found, $t(114.08) = 3.73$, $p < .001$, Hedges' $g = 0.50$, 95% CI [0.24, 0.99], indicating that those in the combined explication and falsification conditions mentioned significantly more exonerating evidence than did those in the control condition.

Table 2

Descriptive Statistics for the Number of Exonerating Pieces of Evidence Mentioned in Each of the Conditions

Condition	*n*	*M*	*SD*
Justification	45	1.31	1.84
Explication	40	1.90	1.77
Falsification	45	1.76	1.92
Control	43	0.79	1.26

5.3.2. Scenarios Considered (H2)

An overview of the number of scenarios considered per condition and scenario type can be found in Table 3. A one-way between-groups ANOVA was conducted to

compare the total number of scenarios that were mentioned in the written decisions across conditions. The analysis showed a significant effect of condition, $F(3, 169) = 3.26$, $p = .023$, $\eta_p^2 = .055$, 90% CI [.004, .106]. A planned contrast was conducted to compare the justification condition to the combined explication and falsification conditions. No significant difference was found, $t(169) = 1.82$, $p = .070$, Hedges' $g = -0.31$, 95% CI [-0.04, 0.69]. Another planned contrast was conducted to compare the control condition to the three experimental conditions combined. A significant difference was found, $t(169) = 2.50$, $p = .013$, Hedges' $g = -0.44$, 95% CI [-0.79,-0.10], indicating that the control conditions included significantly fewer scenarios than the other three conditions. Hence, Hypothesis 2 was partially supported.

An exploratory analysis was also conducted to test whether the number of perpetrators considered by the participants differed across conditions using a one-way independent measures ANOVA. The number of perpetrators was calculated by adding a value of 1 for each mentioned scenario involving a unique perpetrator. No significant difference was found, $F(3, 169) = 2.35$, $p = .074$, $\eta_p^2 = .040$, 90% CI [.000, .084].

Table 3

Number of Scenarios Considered per Condition and Type of Scenario

Condition	*n*	Mean number of scenarios (*SD*)			
		Emma	Mistress	Other	Total
Justification	45	1.00 (0.43)	0.42 (0.50)	0.29 (0.46)	1.71 (0.84)
Explication	40	1.00 (0.55)	0.70(0.69)	0.28 (0.45)	1.98 (0.97)
Falsification	45	1.07 (0.45)	0.80 (0.59)	0.20 (0.40)	2.07 (1.05)
Control	43	1.00 (0.31)	0.30 (0.46)	0.21 (0.41)	1.51 (0.80)

5.3.3. Perception of Guilt (H3)

5.3.3.1. Conviction Rates (H3a and H3b). A Pearson's chi-square analysis was used to determine whether the groups differed in their decisions on whether or not to convict the suspect. Although participants in the control condition

were more likely to convict the main suspect (61.7%) than were participants in the justification (46.8%), explication (48.8%), and falsification (47.8%) conditions, there was no significant difference between the conditions, $\chi^2(3) = 4.33$, $p = .228$, $V = .123$. Hence, Hypotheses 3a and 3b were not supported.

5.3.3.2. Likelihood of Guilt (H3c and H3d). A one-way between-groups ANOVA was conducted to compare participants' ratings of the likelihood that Emma killed James across the conditions (see Table 4). No significant difference was found, $F(3, 169) = 1.22$, $p = .305$, $\eta_p^2 = .021$, 90% CI [.000, .054]. Thus, Hypotheses H3c and H3d were not supported.

Table 4

Likelihood of Guilt Ratings Across Conditions

Condition	*n*	*M*	*SD*
Justification	45	68.53	23.99
Explication	40	66.58	23.11
Falsification	45	66.40	20.48
Control	43	74.23	19.06

5.3.4. Mention of Exonerating Evidence and Conviction Rates (H4)

A point-biserial correlation coefficient had been planned to determine whether the amount of evidence coded as exonerating in the decisions was significantly associated with the binary measure of whether or not the participant would convict the main suspect. However, the Shapiro-Wilk test showed that the assumption of normal distribution was violated for the amount of exonerating evidence mentioned, $W(173) = .798$, $p < .001$. Therefore, a binary logistic regression was conducted instead. The amount of exonerating evidence mentioned by the participant was found to be a significant predictor of their decision on guilt, $\chi^2(1) = 50.35$, $p < .001$, $OR = 2.18$, 95% CI [1.67, 2.85], showing that participants who mentioned more exonerating evidence were more likely to acquit the suspect. Thus, Hypothesis 4 was supported.

5.4. Discussion

In the current study, we aimed to determine whether detailed instructions to account for a legal decision influenced the evidence and scenarios considered by participants. Participants received instructions that were either based on the Dutch (justification) or German (explication) Code of Criminal Procedure, based on the principle of falsification, or conveyed only general instructions to account for their decision (control). Although there was a significant difference between conditions for the amount of exonerating evidence they mentioned in their decisions, there was no significant difference between the justification condition compared to the explication and falsification condition. There was also no significant difference between the justification and control condition, contrary to our expectations. The expectations were based on the idea that the focus on supporting the chosen scenario would result in less consideration of the exonerating evidence (e.g., O'Brien, 2009). Based on the findings, there appears to be a lack of influence from the justification instruction on the use of exonerating evidence. On the other hand, an exploratory analysis showed the combined explication and falsification conditions used significantly more exonerating evidence than the control condition. As the justification condition did not differ from the control condition, while the explication and falsification condition did, it seems the instructions did have a significant effect on the consideration of exonerating evidence, as was expected on the basis of the accountability literature (Lerner & Tetlock, 2003; Tetlock, 1983). The observed pattern could be due to the same mechanism underlying our hypothesis, namely that the justification condition did not include the consideration of alternative scenarios, nor did it encourage the consideration of evidence not supporting the decision (O'Brien, 2009; Van Koppen & Mackor, 2019). Therefore, we had expected the justification condition to be biased towards the guilt of the suspect and the incriminating evidence. Although the difference was not significant between the justification and the other conditions, the fact that it did not differ from the control condition, while the explication and falsification conditions did differ from the control

condition, suggests that participants in the justification condition may indeed have been primarily focused on the guilt of the suspect. Subsequently, the specific request to include evidence beyond that which supports the decision, as in the German but not in the Dutch Code of Criminal Procedure, seems to have resulted in the consideration of more exonerating evidence.

It was expected that the justification condition would consider fewer scenarios than the explication or falsification condition, and that the control condition would consider fewer scenarios than the three experimental conditions. This hypothesis was based on the fact that the instructions in the justification condition did not articulate the need to consider other scenarios (Mevis, 2019) and the fact that the control instructions did not mention alternative scenarios at all. The hypothesis was partially supported, as the control condition did include fewer scenarios than the justification, explication, and falsification conditions. However, the justification condition did not differ from the explication and falsification conditions. Thus, while the detailed instructions in the experimental conditions do seem to have increased the consideration of alternative scenarios, the specific emphasis on alternative scenarios in the explication and falsification conditions was not sufficient to produce a further increase relative to the justification condition. It is therefore unclear whether the explicit instruction to consider alternative scenarios contributes to the use of alternative scenarios above and beyond the effect of providing detailed accountability instructions.

Contrary to our expectations, there was no significant difference in the perception of guilt, in either conviction rates or ratings of likelihood of guilt, between the different conditions. However, we did find, as predicted, that participants who mentioned more exonerating evidence were more likely to acquit the defendant. A possible interpretation of these findings is that, while accountability instructions effectively influenced the consideration of exonerating evidence, the effect was not sufficiently large for it to carry over and influence the global perception of the case. According to that reasoning, the consideration of exonerating evidence can be considered a mediator, which

causally proceeds the outcome variables (i.e., guilt perception and conviction decision). On methodological grounds, it is plausible that a manipulation exerts a stronger influence on the more proximal outcome (i.e., the mediator) than on the more distal outcome (i.e., the dependent variable). With limited statistical power (as in the current study) the proximal effect may be captured, whereas the distal effect goes undetected. Future research should investigate the viability of such causal process models in the context of legal decision-making.

Another possible explanation for the latter finding is that participants in the control and justification conditions, despite not mentioning it in their reasoned decision to the same extent as those in the explication and falsification conditions, did take the exonerating evidence into consideration equally when making their decision. That account would be consistent with research showing that accountability may influence *how* the reasoning occurs rather than *what* the reasoning includes (Hall et al., 2015). It is also in line with the notion that written decisions may contain the evidence that the decision rests on, rather than the evidence that was actually considered in order to come to the decision (Reijntjes & Reijntjes-Wendenburg, 2018). In actual practice, however, the written decision also serves important communicative purposes (Verbaan, 2016). According to the European Court of Human Rights (2019), the reasoned decision is used to show the different parties that they have been heard, which should help them accept the decision that has been made. The reasoned decision is also important to enable parties to use their right to appeal. For those purposes, a discrepancy between what was considered and what is written down could result in a lack of understanding by the parties, and possibly impede the process of appeal. Therefore, decisions made under explication and falsification instructions seem to better serve some of the purposes of a reasoned decision.

In the current study, the communicative functions of the decision were not made explicit to the participants. Instead, they were told that their decision would be evaluated by a panel of experienced judges in terms of how well they followed the instructions. In previous studies on accountability, it has been

suggested that the effect of accountability on decision-making is due to wanting to be viewed positively by others and to avoid receiving criticism (Simonson & Nye, 1992). The effect of accountability we observed may therefore differ from the effects in practice, where judicial reasoned decisions are evaluated, or observed, by several parties, including a court of appeal as well as the defence (Verbaan, 2016). Furthermore, personal consequences, such as receiving criticism, of the reasoned decision were also not included in the current study, although they are likely to affect judges in practice. Further research could determine whether specifying the audiences in accordance with the audience for real life judges increases the influence of accountability on the decision-making process.

5.4.1. Limitations and Future Research

It should be taken into consideration that the current study cannot ascertain whether the need to account for a decision could counter the influence of a prior belief, as expressed in confirmation bias (e.g., Kassin et al., 2013; Nickerson, 1998). We intentionally did not measure the initial perception of guilt prior to participants writing their decision, as stating a hypothesis in itself can cause a preference for information supporting that hypothesis (O'Brien, 2009). We therefore anticipated that stating a hypothesis might cloud the effect of giving a reasoned decision. However, based on the pretest, the case description used was biased towards the guilt of the suspect, which based on previous research (e.g. Rassin et al., 2010; Eerland & Rassin, 2010; Ask et al., 2008) was expected to result in a bias towards incriminating evidence. As indicated by the conviction rates, a strong guilt bias was not observed, suggesting that the need to explain the decision may have countered the influence of the biased initial information across all conditions (including the control condition). That observation supports Gommer's (2007) argument that the requirement for an explanation in itself serves as a countermeasure against the potential influence of bias.

Another limitation of the study is that we cannot be sure how the written decision provided by participants corresponds to their actual consideration

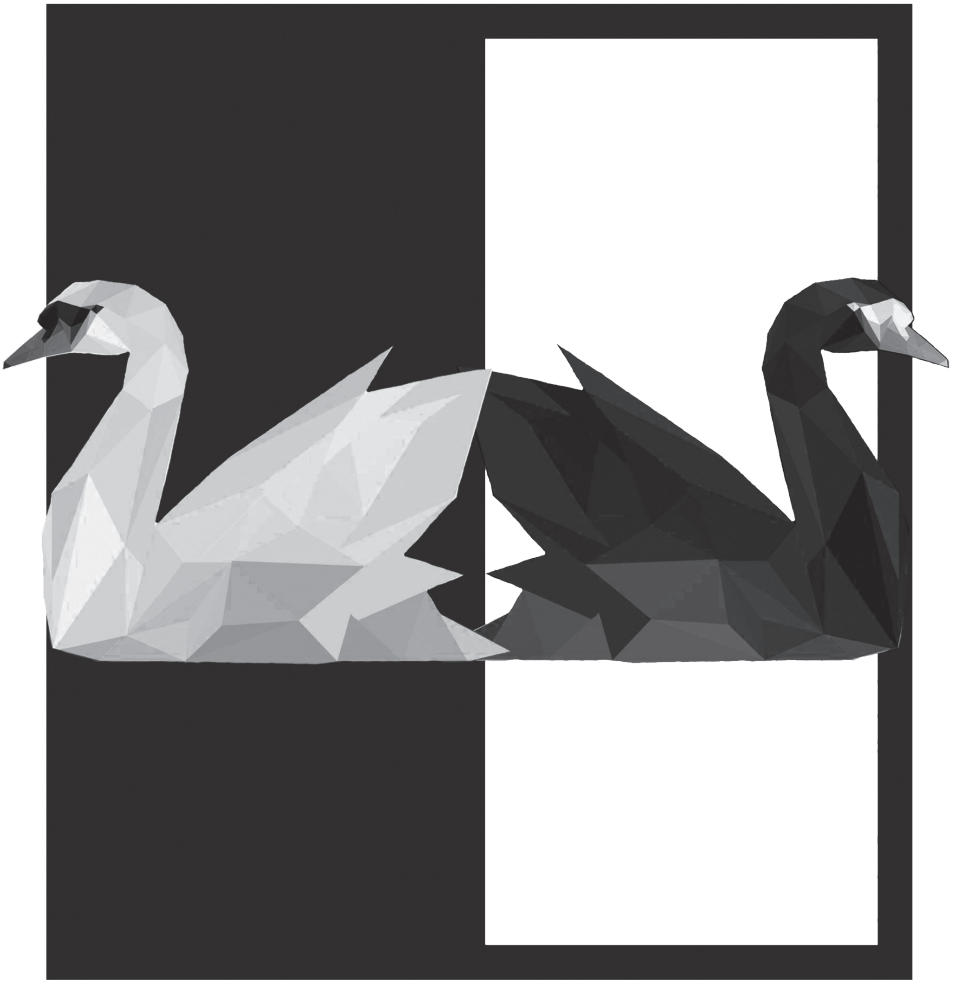
of the evidence. Although their interpretation of the evidence was coded, it is particularly difficult to determine what weight participants assigned to the evidence. The weighing of evidence is also not regulated by legal rules (Anderson et al., 2005). Furthermore, as participants were specifically asked to include certain elements in their reasoned decision, dependent on their condition, the extent to which the mention of the evidence means it was actually considered remains unclear. However, as the mention of exonerating evidence was a significant positive predictor of the decision to acquit, we can tentatively conclude that the inclusion of exonerating evidence also means participants attached value to the exonerating evidence. Further research into the weighing of evidence could also contribute to the understanding of reasoned decisions.

A further limitation of the current study is that no condition without instruction was included in the design. Although we intentionally did not include such a condition as it would not be a realistic representation of judicial decisions in practice, it also limits the conclusions that can be drawn on the basis of our findings. Despite the lack of ecological validity, a condition without instruction would have allowed us to compare the need to account for a decision to not having to explain a decision at all. It would therefore be advisable, in future research, to include a condition which requires no explanation at all, in addition to the detailed instructions researched in the current study.

Lastly, a limitation of the current study relates to statistical power and precision. First, the sample was somewhat smaller than we had initially aimed for. Second, our power analysis was based on the detection of a medium-sized effect. Therefore, the current study was acceptably powered to detect only a medium-to-large effect size. As a consequence, we cannot exclude the existence of smaller effects that may nonetheless be of practical relevance. Furthermore, the observed effect sizes in the current study were not precisely estimated (as indicated by wide confidence intervals), which means that we may have over- or underestimated the true underlying relationships. The accumulation of further empirical evidence is necessary to establish the robustness and validity of our statistical conclusions.

5.4.2. Conclusion

Overall, the findings of the current study have positive implications for the requirement imposed on judges to explain their decision. Our findings indicate that variations in the instructions as to how (mock) judges should explain their decisions can influence the type of evidence considered when making the decisions. We did not find evidence, however, that instructions focusing primarily on incriminating evidence (as dictated by the Dutch Code of Criminal Procedure) negatively affect the final decision on guilt. The lack of statistical power, however, prevents us from concluding that the effectiveness of instructions based on the principles of justification, explication, and falsification does not differ. Although the explication and falsification instructions led to an increased use of exonerating evidence compared to control instructions, this did not translate into differences in the final decision on guilt. This finding suggests that a key component of the German Code of Criminal Procedure—the requirement of judges to explicate their decision and consider alternative scenarios—may improve the transparency and thoroughness of reasoned judgements.



CHAPTER 6

TEST OF THE ANALYSIS OF COMPETING HYPOTHESES IN LEGAL DECISION-MAKING

An adapted version of this chapter has been published online as:

Maegherman, E., Ask, K., Horselenberg, R., & Van Koppen, P. J. (2020) Test of the analysis of competing hypotheses in legal decision-making. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.3738

Abstract

The analysis of competing hypotheses (ACH) has been suggested to be a method that can protect against confirmation bias in the context of intelligence analysis. In the current study, we aimed to determine whether ACH could counter confirmation bias in the reasoning with evidence in the context of criminal law proceedings. Law students (N = 191) received information about the ACH method or general information about biases. They were given a case vignette with a main suspect and a list of 24 questions, six of which they could ask about the case. Half of the questions related to incriminating information, whereas the other half related to exonerating information. Contrary to our expectations, participants in both conditions favoured questions relating to exonerating information and rated the exonerating evidence as being more important for their decision. Despite the lack of bias observed, it seems participants failed to properly apply the ACH method.

6.1. Introduction

In criminal law proceedings, the judges or juries have to determine whether suspects are guilty of the offence they are accused of. Such a decision requires the decision maker (hereafter referred to as judge) to reason from the available evidence to a final judgement. To do so, the judge has to decide what has been proven to be true. That process can be compared to testing hypotheses about what happened; the guilty scenario presented by the prosecution, and possibly the alternative scenario of innocence most likely presented by the defence. These can then be tested using the available evidence. Several theories exist about how a decision is made in legal proceedings (e.g., Simon, 2004; Pennington & Hastie; 1988; Wagenaar et al., 1993; Van Koppen & Mackor, 2019). The reasoning processes undertaken by decision makers in criminal law proceedings are inherently cognitive processes, and subsequently, can be vulnerable to cognitive biases. Such cognitive biases can also lead to miscarriages of justice (Bandes, 2006; Martin, 2002). In the present study, we investigated whether a structured analysis technique taken from the field of intelligence analysis, namely the analysis of competing hypotheses, can reduce the effect of cognitive biases.

An important aspect of determining whether something is true, addressed by Popper (1959), is the need for attempted falsification rather than merely confirmation of the theory to be tested. Falsification refers to searching for facts that disconfirm the existing hypothesis. Popper (1959) pointed out that, even with evidence supporting the theory, a theory can only be maintained until evidence that disconfirms it is found. The fact that a theory has not been disproven yet does not mean it must be true. In short, when trying to determine whether something is true, one must look not only at evidence supporting that, but must foremost try to find evidence that falsifies the rule. Failed serious attempts at falsification increase the likelihood of the hypothesis being true (Crombag et al., 2006).

6.1.1. Confirmation Bias

Although attempting falsification might seem like a simple process, it is complicated by a common cognitive bias, namely confirmation bias (Wason, 1968). When confirmation bias occurs, one focuses excessively on finding evidence that confirms an initial hypothesis, while paying disproportionally little attention to evidence that contradicts that hypothesis (Nickerson, 1998). That hypothesis can then be maintained, even when faced with contradicting evidence, which is known as *belief perseverance*. Belief perseverance has also been observed in some prosecutors who, in exoneration cases, have maintained that the convicted individual was guilty despite overwhelming evidence to the contrary (Burke, 2007). The biased interpretation and selection of evidence which constitute confirmation bias consequently help maintain the belief. For instance, confirmation bias could affect the interpretation of evidence in such a way that it becomes more supportive of the existing hypothesis (Kassin et al., 2013). It has been argued that confirmation bias is particularly relevant in legal proceedings (Findley & Scott, 2006). In addition, in some jurisdictions (e.g., the Netherlands), the case file presented to judges is likely to be biased towards the guilt of the suspect, which has been argued to make judges prone to confirmation bias (Crombag, 2017).

Schünemann and Bandilla (1983) tested the influence of incriminating information presented to judges prior to trial on judges' decisions. One group received pre-trial information that was mainly incriminating for the suspect, while the other group received information that was less incriminating prior to trial. All judges then received identical records of the subsequent proceedings. In the first group, 82% of judges chose to convict the suspect, whereas only 53% of judges in the second group chose to convict the suspect. Based on these findings, it appears that judges' decisions are strongly affected by their initial belief. As discussed, according to Popper's argumentation (1959), judges should also try to falsify their hypothesis. According to Wagenaar et al. (1993), the simplest way to try to falsify the criminal charge would be to investigate what happened according to the suspect. However, in line with research on confirmation bias

(Wason, 1968; Jonas et al., 2001; Ask & Granhag, 2005), one would expect that there will be a preference for evidence that confirms the initial hypothesis, presumably the charge, and a subsequent failure to attempt falsification of that hypothesis.

In light of the potential detrimental effect of confirmation bias, it is paramount to examine the relation between confirmation bias and the process of falsification. Mynatt and colleagues (1977) used a computer-controlled environment to this end. Participants were asked to formulate a hypothesis, after which they were asked to choose between an environment where they could make observations that would likely confirm their hypothesis, or an environment where they could test alternative hypotheses. Participants showed a strong preference for the former environment. Participants who did receive falsifying information used that information to reject incorrect hypotheses.

Other researchers have found that participants presented with a criminal case file show a preference for evidence that supports their existing beliefs. For instance, Rassin et al. (2010) presented students with a case file and asked them to decide whether the suspect was innocent or guilty. Participants could then choose from 20 further investigative measures, half of which were directed at obtaining further evidence against the suspect, and half which were aimed at obtaining exonerating information. The initial decision was found to predict the further information sought, with participants choosing investigative measures fitting their initial decision. Similar findings were presented by O'Brien (2009), who found that participants who formulated a hypothesis about the likely culprit in a mock crime investigation showed more confirmation bias than those who did not formulate a hypothesis or who formulated both a hypothesis and a counterhypothesis. Interestingly, participants who were asked to write down three alternative suspects did not show less bias than those who only identified one primary suspect. Therefore, whether and how the formulation of multiple alternative hypotheses can reduce bias requires further investigation.

The consideration of alternative scenarios in legal decision-making has also been investigated by Rassin (2018). In his study, some participants were

presented with an alternative scenario in a criminal case whereas others were additionally instructed to indicate how well each piece of evidence fitted with the main and the alternative hypothesis, respectively. They used pen and paper to do so, which means it was an active task rather than simply having them consider the evidence. Participants were asked to rate police findings and to decide whether or not to convict the main suspect. Rassin (2018) found that the pen-and-paper tool counteracted an excessive focus on evidence confirming the main hypothesis in both lay people and criminal justice professionals. The pen-and-paper tool group also had a lower conviction rate than the alternative scenario group. When combined with the findings by O'Brien (2009), it therefore seems that only formulating alternative scenarios may not be sufficient; one should actively consider the alternative scenarios in order to counteract the influence of confirmation bias.

6.1.2. Analysis of Competing Hypotheses

A potential tool to structure the consideration of alternative scenarios is the Analysis of Competing Hypotheses (ACH). ACH is a procedure which was originally designed to help intelligence analysts avoid common pitfalls (Heuer, 1999). It requires careful weighing of alternative explanations. These are compared against each other, thus preventing the analyst from deciding on the first solution that seems satisfactory. One of the pitfalls targeted by ACH is the concern that focusing on trying to confirm one hypothesis may cause analysts to fail to recognise that much of the evidence is also consistent with other explanations or conclusions that have not yet been refuted (Heuer, 1999). It is evident that that problem shows similarities with confirmation bias. It would therefore be useful to determine whether the ACH method can also be used in other areas where confirmation bias is likely to play a role, such as decision-making by judges.

The ACH method consists of eight steps, as summarised below on the basis of Heuer´s (1999) description. The first step requires the identification of potential hypotheses; that can be accomplished in collaboration with colleagues who hold

differing views. The second step is to make a list of significant evidence, noting not only the presence of evidence but also the absence. For each hypothesis, it should be considered what you would expect to see, and not see, if it were true. During the third step, a matrix is prepared with hypotheses across the top, and evidence down the side. The matrix is then used to analyse how each piece of evidence relates to each hypothesis. In that step, the analyst works across the matrix, examining how consistent each piece of evidence is with each of the hypotheses. In the corresponding cell of the matrix, it should be noted whether the evidence is consistent, inconsistent or irrelevant to the hypothesis. The matrix aids the determination of the diagnosticity of the evidence. For instance, if a piece of evidence is consistent with all hypotheses, its diagnostic value is very low to non-existent. In the fourth step, the matrix should be refined. It may be possible that some hypotheses should be added, or more distinctions should be made to include other alternatives. Evidence which has no diagnostic value should be removed from the matrix.

During the fifth step, tentative conclusions can be drawn about the likelihood of each hypothesis. It is important to try to disprove hypotheses rather than trying to prove them. The hypothesis which has the fewest inconsistent pieces of evidences is probably the most likely. However, that does not mean that the hypothesis with the most consistent pieces of evidence is the most likely. The matrix only serves as a tool that causes giving more consideration and more analysis of hypotheses that would otherwise have been thought to be unlikely. During the sixth step, it should be considered how sensitive the conclusion is to a few critical items of evidence, and what the consequences would be if that evidence were misleading or subject to a different interpretation. Additional investigations may be needed. Subsequently, conclusions of the analysis are drawn during the seventh step. The relative likelihood of all hypotheses should be discussed. It is important to proceed by eliminating hypotheses, rather than by confirming them. During the final step, analysists should identify possible future information that may change the outcome of the analysis (Heuer, 1999).

Because it was initially developed for use in intelligence settings, it is worth studying whether the ACH procedure can also be used in the consideration of evidence in judicial proceedings. If found to be effective, the procedure would then offer a systematic approach to decision-making, which could reduce the influence of biases. To date, ACH has not yet been the subject of many studies (Puvatinghal & Hantula, 2012). In one study, Lehner and colleagues (2008) found that ACH mitigated confirmation bias for participants who did not have experience working in intelligence analysis. That manifested itself in weighing of the evidence rather than in evidence interpretation. Dhami et al. (2019) found mixed evidence for the effectiveness of ACH in reducing confirmation bias in trained analysts instructed to use the ACH method, as well as a lack of adherence to the method. The current study examined whether informing people about the ACH method can reduce the influence of an initial belief by increasing the attempts made to falsify the original hypothesis in the context of a criminal case.

6.1.3. The Current Study

Based on existing theory and previous research, in the current study, we examined the relation between an initial belief, attempts at falsification, and the final decision using a mock criminal case. Furthermore, it was tested whether training in ACH can affect the extent of attempted falsification. In order to do so, participants were either trained in the ACH procedure, or merely received general information on biases. Participants were then provided with a description of a case, in which a guilty scenario for the main suspect was obvious, and an alternative scenario was subtly implied. After participants had indicated their initial belief about the case they were given the opportunity to receive answers to a number of investigative questions. The questions related to information with potential to either confirm or disconfirm their initial belief. Answers to the disconfirming questions could potentially either support the alternative scenario or give evidence that contradicted the scenario that the main suspect was guilty (e.g., providing an alibi, finding traces of an alternative suspect). After obtaining

additional information based on the selected questions, participants were again asked to indicate their belief about the case by, for instance, rating the likelihood that the main suspect was guilty.

We expected that those who were instructed in the ACH procedure would show more attempts at falsification than those in the control condition, by choosing more questions that could potentially disconfirm the main suspect's guilt (H1). Those in the ACH condition were subsequently expected to have a lower final rating of likelihood of the suspect's guilt than those in the control condition (H2).

6.2. Methods

6.2.1. Participants

Participants were recruited through advertisement at various law faculties in the Netherlands and Belgium. Of the original sample of 361 participants, 170 participants were excluded for reasons related to failing the attention or control checks, or having a background in something other than law (for a detailed overview of the exclusion of participants, see the Supplemental Material). Participants could choose to take the survey in Dutch or in English. The final sample included 191 participants, of which 154 chose to take the survey in Dutch. Participants' average age was 22.8 years (SD = 3.5). The majority of the participants were female (68.6%). Out of the final sample, 71.7% completed the survey online. A sensitivity power analysis based on the final sample size, a power of .80, and a .05 significance level was conducted using the G*power software (v3.1; Faul et al., 2007). According to the power analysis, the collected sample size was acceptably powered to detect an effect size of d = .36. Ethical approval had been obtained from the Ethical Committee at the Faculty of Psychology and Neuroscience at Maastricht University.

6.2.2. Materials

6.2.2.1. ACH and Biases Information. In the ACH condition, participants were given a short explanation of the ACH procedure. The explanation focused

on producing several hypotheses and using the matrix, as these are thought to be the key components of the method relevant to decision-making. In addition, participants were given a short description of confirmation bias. To ensure that the only difference between the conditions was the explanation of ACH, participants in the control group also received the information on confirmation bias (Appendix F).

6.2.2.2. Case Description. A short fictional case description about a criminal investigation was used. In the case, a man named Jasper Kostal had been murdered. The main suspect was his wife, Sabine. The case contained information that appeared incriminating for Sabine (e.g., she was covered in blood, her fingerprints were on the knife). However, an alternative scenario was also subtly implied in the case description. Specifically, there was evidence that Jasper had been having an affair (Appendix F). The case was pretested by asking participants to rate the likelihood that Sabine was guilty of killing Jasper on a scale from 0 (*not at all likely*) to 100 (*very likely*). As the case had to be perceived as incriminating against Sabine, it was adapted in response to the pretest ratings. The case was also edited for authenticity by an experienced court clerk. In the final pretest (N = 20), the average rating for the likelihood that Sabine was guilty of killing Jasper was 73.4 (SD = 19.3).

6.2.2.3. Investigative Questions. Participants were presented with a list of 24 investigative questions (Appendix F). Participants were asked to choose six questions in total, and were returned to the list of questions after having received the answer to each question. To 12 of these questions, the answer was confirming for the obvious scenario of guilt (e.g., is Sabina right-handed?), whereas the other 12 answers were disconfirming for the obvious scenario of guilt (e.g., is there evidence Sabine tried to do CPR on Jasper?). Answers to the disconfirming questions could indicate that the obvious scenario could not be true, provided alternative explanations for the evidence found, or provided support for the alternative scenario. In order to ensure that participants perceived the questions to be confirming or disconfirming for their hypothesis as intended, participants

were asked to report why they had chosen to ask each question using open-ended answers. Although we had initially planned to use the questions selected by participants as a measure of falsification attempts, it soon became clear during coding that this was not feasible. Based on participants' responses, their interpretation of the question did not always match the classification we had given it. For example, for the question asking whether Sabine is right-handed, some participants emphasized that it would have been exonerating for Sabine if she had turned out to be left-handed, whereas others emphasized that Sabine's right-handedness fits with the hypothesis that she was the perpetrator. It thus became clear that the selected questions could not easily be classified as confirming or disconfirming without knowing what outcome the participant had anticipated. Unfortunately, participants' open-ended answers did not always offer insight into their expectations, as several answers were of the type "it's important to know", "I was wondering", or even "just checking". It was therefore decided to use an alternative measure of falsification (see below).

6.2.2.4. Importance Ratings. At the end of the study, participants were asked to rate the importance for their final decision on guilt of each piece of evidence that they had received as answers to the investigative questions, using a scale from 0 (*not at all important*) to 100 (*very important*). As these ratings related to the answers to the investigative questions, rather than to the questions themselves, the interpretation of these ratings was much more straightforward and thus more suitable than question selection as a measure of preference for falsifying information. In testing H1, it was therefore expected that participants in the ACH condition would have a higher average importance rating for the evidence that falsified the main suspect's guilt than would those in the control condition.

6.2.3. Procedure

Initially, the study was administered in person as we considered also using the side notes made by participants. Some of these participants came to the lab, whereas others were tested in a classroom setting. However, data collection

in person proved to be extremely difficult, and the notes participants made had little additional value. We noticed that participants were often not making notes, or merely writing down a few keywords of the case rather than using a structured approach as instructed by the ACH instruction. Data collection was therefore continued online. The survey platform Qualtrics was used. Several questions were added to ensure the quality of the gathered data. These included four control questions about the case, two control questions about the ACH instruction, and three attention checks throughout the procedure.

After providing informed consent, participants either read only general information about biases (control condition) or read general information about biases *and* received an explanation of the ACH procedure (ACH condition). In the online study, participants in the ACH condition then also received control questions related to the ACH manipulation (e.g., a true or false question asking whether the main focus was on evidence supporting the hypothesis). They were told how many questions they answered correctly and were then given the option to revisit the ACH instruction. Participants were then given the case file. After having read the case file, participants were asked to rate the likelihood that Sabine was guilty (0 = *not at all likely*, 100 = *very likely)*, whether or not they would convict her for the murder of her husband[10], and how confident they felt about their conviction decision (0 = *not at all confident*, 100 = *very confident).* These answers were used as a measurement of their initial belief. Furthermore, participants were asked to write down their hypothesis or hypotheses about what happened. For these, the number of scenarios formulated were counted, as well as the perpetrators named in the scenarios.[11] Participants were then given the list of confirming and disconfirming investigative questions.

All participants selected six investigative questions and received answers to those questions. Participants were again asked to rate the likelihood of the

10 The conviction ratings were analysed and can be found in the Supplemental Materials (Appendix G).

11 The exploratory analysis of the formulated scenarios can be found in the Supplemental Materials (Appendix G).

suspect being guilty, whether or not they would convict the suspect, and how confident they felt about their conviction decision. Finally, participants were then asked to rate the importance for their final decision on guilt of the evidence they had received as answers to the investigative questions.

Participants took an average of 1 hour and 47 minutes to complete the study (*SD* = 8 hours and 58 minutes). The median response time was 28 minutes and 7 seconds.[12]

6.3. Results

6.3.1. Language

Participants could choose to complete the survey in either Dutch or in English. The Dutch and English groups did not differ for the measures of likelihood of guilt, the number of confirming or disconfirming questions selected, or importance ratings. Language also did not interact with condition for any of the above measures (detailed results for the analyses involving language can be found in the Supplemental Materials; Appendix G). It was therefore decided that language was not a confounding factor and it was not included in subsequent analyses. The data from the Dutch and the English groups were combined for all further analyses.

6.3.2. Question Selection

Based on Shapiro-Wilk tests, it was concluded that the data for the dependent variables were not normally distributed (detailed outcomes for the analyses can be found in the Supplemental Materials in Appendix C). It was therefore decided to conduct non-parametric test for the main analyses.

In order to determine whether those instructed in ACH showed a stronger preference for the questions which were potentially disconfirming for the suspect's guilt, the questions chosen were compared between the ACH and the

12 We suspect the large mean and standard deviation can be explained by participants not closing the tab immediately after they completed the survey. The longest response time was 4.5 days.

control condition using Mann-Whitney U tests. Participants could choose a total of six questions. The number of disconfirming and confirming questions was calculated for each participant. Participants in the ACH condition (*Mdn* = 2.0) and the control condition (*Mdn* = 2.0) did not differ significantly in the number of selected questions that were confirming of guilt, *U* = 4289.50, *p* = .476, *r* = .052. The ACH group (*Mdn* = 4.0) also did not differ from the control group for the number of disconfirming questions chosen (*Mdn* = 4.0), *U* = 4289.50, *p* = .476, *r* = .052. A Wilcoxon signed rank test indicated that participants overall chose significantly more questions that were disconfirming (*Mdn* = 4.0) than questions that were confirming for the suspect's guilt (*Mdn* = 2.0), *z* = -9.05, *p* < .001, *r* = .463.

6.3.3. Importance Ratings

The importance ratings for the exonerating evidence was averaged for each participant. The same was done for the importance ratings for the incriminating evidence. Participants in the ACH condition (*Mdn* = 74.0) did not differ significantly from participants in the control condition (*Mdn* = 71.0) in their importance ratings for the exonerating evidence, *U* = 3896.50, *p* = .202, *r* = .093. There was also no significant difference between the ACH group (*Mdn* = 56.3) and the control group (*Mdn* = 54.0) in the average rating for incriminating evidence, *U*= 3623.50, *p* = .830, *r* = .016. Overall, participants rated the exonerating evidence as significantly more important (*Mdn* = 72.7) than the incriminating evidence (*Mdn*= 55.3), *z* = -7.13, *p* < .001, *r* = .376. As there was no evidence that participants in the ACH condition showed a stronger preference for disconfirming evidence than did control participants, we found no support for H1.

6.3.4. Ratings of Guilt

Participants were asked to rate the likelihood of the main suspect, Sabine, being guilty immediately after reading the case (Time 1) and again after they had received answers to the investigative questions (Time 2). Participants in the ACH condition (*Mdn* = 70.0) did not differ significantly from participants in the control

condition (*Mdn* = 69.0) for the rating of likelihood of Sabine being guilty at Time 1, *U* = 4281.00, *p* = .479, *r* = .051. For the likelihood rating at Time 2, participants in the ACH condition (*Mdn*= 40.0) also did not differ significantly from participants in the control condition (*Mdn*= 45.0), *U* = 4381.00, *p* = .657, *r* = .032. This finding indicates that there was no support for H2. Overall, the ratings for likelihood of guilt decreased significantly between Time 1 (*Mdn* = 70.0) and Time 2 (*Mdn* = 40.0), *z* = -10.27, *p* < .001, *r* = .526.

6.3.5. Compliance with the ACH Instructions

Participants in the ACH condition were asked whether they had made an ACH matrix, how helpful they found it, and how difficult they found it after giving the importance ratings. Of the participants in the ACH condition, 38% reported having used the matrix. Participants are therefore thought not to have complied with the ACH instructions. The average helpfulness rating for those who reported using the matrix was 57.1 (*SD* = 18.8) on a scale from 0 (*not helpful at all*) to 100 (*very helpful)*. The average difficulty rating was 46.3 (*SD* = 28.5) on a scale from 0 (*not difficult at all*) to 100 (*very difficult*). An independent ANOVA was used to test for a difference between those who made an ACH matrix, those who did not make an ACH matrix, and those in the control conditions. No significant difference was found for the question selection, for any of the ratings of likelihood of guilt, nor for the importance ratings of either the exonerating or the incriminating evidence. Detailed outcomes of the analysis can be found in the Supplemental Materials (Appendix C).

6.4. Discussion

The current study was intended to evaluate whether ACH could protect against confirmation bias in the context of a criminal case. This question proved difficult to answer as, contrary to our expectations, participants in the current study did not show confirmation bias in either condition. Participants in both conditions seemed to favour questions that were disconfirming for the

guilt of the main suspect in the case. Similarly, they also rated the exonerating evidence as more important for their decision than the incriminating evidence. Their perception of guilt of the main suspect also decreased between time 1 and time 2, as did the conviction rates for the suspect. Moreover, participants in the ACH condition did not differ significantly from the participants in the control condition in their initial ratings of likelihood of guilt. Participants in both groups had a relatively high rating of guilt, and the majority of participants chose to convict Sabine after only having read the case file. Subsequently, they also did not differ between conditions for the questions they chose to ask, or in their importance rating of the exonerating and incriminating evidence. Therefore, it appears there was no difference between those who received the ACH instructions and those who did not receive the ACH instructions in terms of perception of guilt, nor in their search for further information.

A significant difference was, however, found for the ratings of guilt and conviction rates at Time 2 compared to Time 1. The perception of guilt decreased after participants received answers to the chosen questions. That finding is in line with the fact that participants also chose more questions that we had originally coded as disconfirming of the suspect's guilt than questions that were considered confirming of the suspect's guilt. Similarly, participants overall rated the exonerating evidence as more important for their decision than the incriminating evidence. The results therefore support the conclusion that participants were not biased towards the scenario of the main suspect and updated their belief in the main suspect's guilt appropriately in response to the exonerating evidence received.

There are several potential explanations for these findings. Firstly, the lack of confirmation bias found in all participants should be considered. Confirmation bias was expected based on the fact that participants received an initial case file that was biased towards guilt of the main suspect (O'Brien, 2009). Even though the case file also contained the suggestion of an alternative scenario, preliminary testing of the material still indicated that it was biased towards the guilt of

Sabine, with an average rating of guilt of 75 (out of 100). Furthermore, the initial ratings of guilt in the main study also supported that the case file initially created a belief in guilt. Therefore, it seems that the lack of confirmation bias was not due to a lack of an initial impression of guilt. It may be so that the manipulation failed to create confirmation bias as, despite the impression of guilt, participants were not sufficiently invested in the case. A lack of investment could also have resulted in less cognitive dissonance experienced in response to contradicting information, and subsequently, no preference for supporting information in order to achieve or maintain consonance (Ask et al, 2011; Festinger, 1957). Obtaining such an investment from participants would likely require a task that allowed them to be more cognitively or emotionally involved.

Nevertheless, a preference for incriminating information following an initial impression of guilt has been observed in several studies (e.g., Ask & Granhag, 2005; O'Brien, 2009; Marksteiner et al., 2011). Rassin (2018) also found that participants showed tunnel vision in response to an incriminating case file. In that study, the effect of tunnel vision was not reduced by the introduction of an alternative scenario, but using a simple pen-and-paper tool to indicate how well each piece of evidence fit with each of the scenarios did reduce tunnel vision. The findings of the current study are thus not in line with previous research, as the initial impression of guilt did not cause a preference for incriminating information. It seems unlikely that the lack of confirmation bias was due to the suggestion of an alternative scenario in the current study, as that has previously been found to be insufficient to protect against confirmation bias (O'Brien, 2009; Rassin, 2018).

Although the lack of confirmation bias and the appropriate interest in the exonerating materials are, of course, encouraging findings, they were also unexpected based on earlier research. One possible explanation that has also been offered in other studies in this thesis is that the lack of confirmation bias might be due to the sample of law students that was used. Law students were chosen as the population for this study due to the context of the material and the

fact that finding real judges to participate in research has proven very difficult. Law students were thought to have an affinity with the reasoning required for the current study, which was not expected from a community sample or psychology students. However, the element that may have been underestimated is that law students are also the lawyers of the future. A possible implication of the findings might therefore be that law students have a keener eye for alternative scenarios, and might have a greater interest in the exonerating evidence, than expected. A similar finding was reported by Rassin et al. (2010), who also used law students as participants. In their study, participants chose mainly exonerating investigations in response to an incriminating case file, although the amount of incriminating investigations increased in response to more evidence and increased severity of the case. One question that remains is whether the critical stance observed in law students in experimental research will ultimately also result in increasingly critical judges in practice, or whether external factors can threaten their openness to alternative explanations.

Another potential explanation for the lack of observed confirmation bias in the current study could be that participants in the control condition were also warned about biases prior to reviewing the case file. However, that seems unlikely to provide a sufficient explanation for the lack of bias when considering the existing evidence for a bias blind spot (Pronin et al., 2002). The bias blind spot has recently also been documented by Kukucka and colleagues (2018) in an international survey of forensic examiners. Examiners showed a tendency to acknowledge bias in other domains than their own, as well as in other examiners but not in themselves. Therefore, it seems unlikely that merely informing participants about biases would have protected against confirmation bias in the current study.

While neither of the groups showed confirmation bias, so that ACH could not counter confirmation bias, there were also exploratory findings indicating that ACH was not properly used by participants. For instance, based on the notes participants made, it was clear that they were mainly making a list of evidence rather than using the ACH matrix as instructed. Secondly, when participants

were asked whether they had used the matrix, the majority responded they had not used the matrix. Of those who did, the average rating for how helpful they found the matrix was only just over the midpoint, whereas the difficulty level they reported was just under the midpoint. Therefore, the intended use of ACH did not seem to be adopted by most participants in the ACH condition.

There are several possible explanations related to the ACH method itself as to why ACH did not work or was not used. One of these is the lack of training that participants received. Participants were given a simplified description of the ACH method, which had been pre-tested for clarity. They were also given an example case on which they could practice. They were then asked two control questions about the instructions. Although participants who answered those questions wrong were excluded, we could not check whether they had actually practiced on the example case. Therefore, the instructions participants received may not have been applied by participants. One potential implication is therefore that ACH may require in-depth training before it can be used at all, let alone effectively. Dhami et al. (2019) also found that the majority of intelligence analysts in their study who had been trained and instructed to use ACH did not follow one or several of the steps prescribed by the technique.

Lastly, ACH has also received criticisms for various reasons in more recent research. For instance, Dhami et al. (2019) pointed out that the criteria used in ACH are vague, which can make the judgement process by the analysts unreliable. It is unclear which criteria should be used to determine whether evidence is consistent or inconsistent with a scenario. They also found that those trained in ACH were less consistent in evidence assessment and in final their conclusions, compared to earlier decisions, than those who had not been trained in ACH. Criticisms of ACH have also been voiced in more general terms about structured analytic techniques, which ACH falls under (Chang et al., 2018). The lack of research supporting the effectiveness of these techniques has also repeatedly been used to question their use in practice (Chang et al., 2018; Dhami et al., 2019). Several variations of the ACH method have been suggested in an

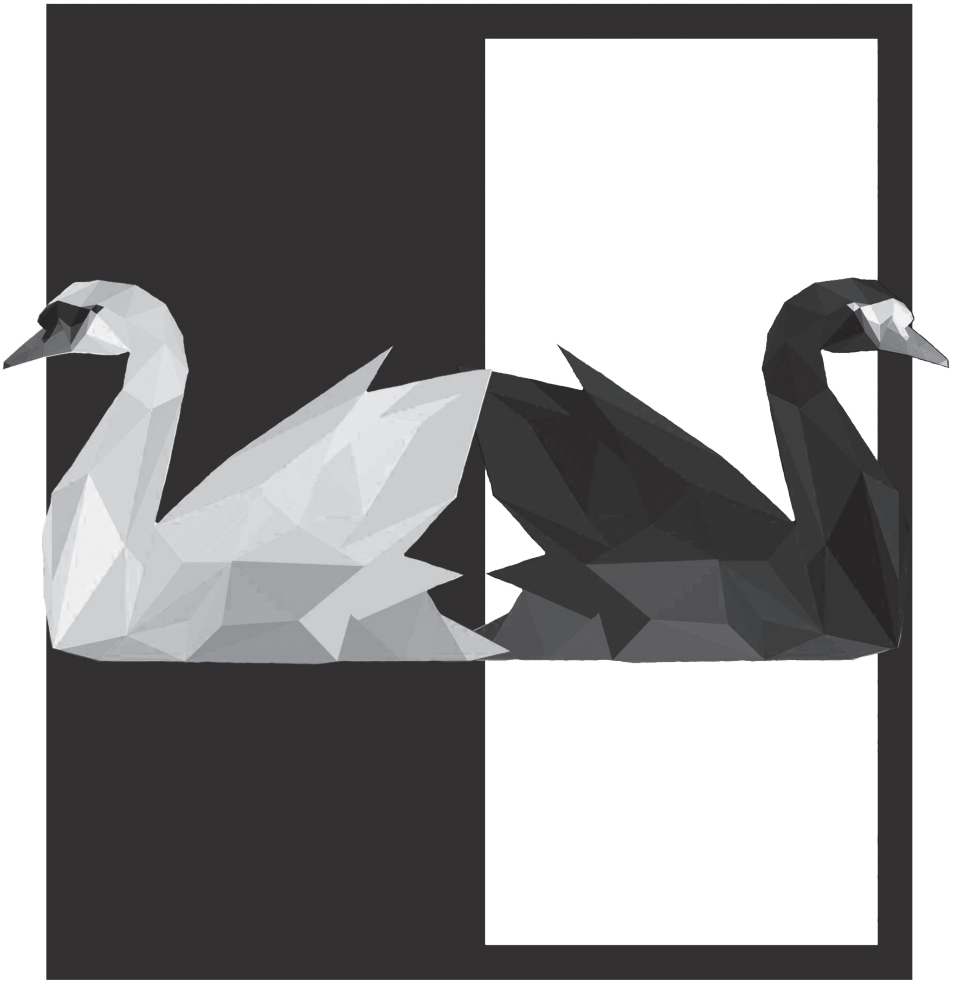
attempt to improve the original ACH method. These include for instance the argumentation-based ACH (Murukannaiah et al., 2015) and a collaborative, web-based, version of ACH (Convertino et al., 2008). However, further research is needed to determine the effectiveness of both the original and adapted ACH methods in a variety of contexts where confirmation bias exists.

Beyond the criticisms of the ACH method, limitations of the current research include the fact that the study was conducted online for a portion of the participants. In order to protect against this limitation, a number of attention checks were included throughout the study. These included directed queries (i.e., answering the question in a specified way) and memory tests (Abbey & Meloy, 2017). Furthermore, participants also had to answer control questions about the ACH instruction and about the case. Participants also had the option to review the case at several points during the study, so there was no limit on their access to the necessary material due to the study being online.

Another limitation of the study is that we did not know what participants' expectations were when they chose an investigative question. In other words, we could only judge what they were testing by inference from the question they asked, which may not be a valid measurement of what they tried to find out. In future research, it would be beneficial to ensure that participants also have to indicate what outcome they are expecting.

6.4.1. Conclusion

Findings of the current study did not show evidence of confirmation bias among participants in the control condition or the ACH condition, thereby preventing us from conclusively testing whether ACH could protect against confirmation bias. Despite that difficulty, it was nevertheless clear that participants did not adequately make use of the ACH method. In order to validly determine whether ACH can prevent against confirmation bias in the context of criminal proceedings, further research is needed, although it may be advisable to use an adaptation of the ACH method.



CHAPTER 7

GENERAL DISCUSSION

7.1 Introduction

When deciding whether or not a suspect is guilty, judges have to weigh the evidence that is available to them in order to come to a decision. Despite extensive regulations on admissibility of evidence, there is almost no formal regulation on how evidence should be evaluated or on the weight that should be attributed to evidence (Anderson et al., 2005). In miscarriages of justice, tunnel vision, or an overreliance on incriminating evidence, has been identified as a contributing factor (e.g., Findley & Scott, 2006; Posthumus, 2005; Sorocham, 2008). In the current thesis, I examined the consideration of exonerating evidence and alternative scenarios in the context of decision on guilt in criminal law proceedings. The approach used was largely based on the theories on legal decision-making and evidence consideration that were outlined in Chapter 1. The focus was on determining how falsification, or the attempt to find evidence against a theory of guilt, happens in practice, how that differs from what is prescribed by theory on legal decision-making, and how it can be facilitated in practice.

The studies presented in Chapters 2 and 3 provide insight into how exonerating evidence is used in practice through a case study and a survey with judges. From both of those studies, it appeared that judges were willing to use falsification and consider alternative scenarios. However, they struggled to do so in practice. In the remainder of the studies, I aimed to experimentally test methods through which falsification could be facilitated.

In the rest of the current chapter, I discuss the findings of the studies presented in this thesis, taking into account the extent to which these findings have implications for practice. I will discuss the findings regarding falsification in legal decision-making in current practice, and possible areas for improvement that have been identified in my research. Therefore, I will also address possibilities for future research that have come to light through the current research.

7.2 Summary of Findings

In this thesis, I examined the use of falsification in the context of evidence consideration in criminal law proceedings. Particularly, I tried to find out whether and how exonerating evidence is currently considered by judges in their decision-making process, and how this related to the concept of falsification (Popper, 1959/2005), and the established phenomenon of confirmation bias (Nickerson, 1998). By analysing a case offered to the Project Reasonable Doubt (Amsterdam Lab for Legal Psychology, n.d.), and by distributing a questionnaire amongst judges and interviewing judges on that theme, I tried to gain insight into how falsification is used in practice, followed by several experimental studies on ways in which falsification could be facilitated.

7.2.1. Falsification in Legal Practice

7.2.1.1. Case Study. In Chapter 2 and Chapter 3, I reported on examining judges' current perspective on falsification through the use of a survey and a case study. Results from the case study showed that judges clearly tried to use alternative scenarios in their decision-making process. However, the manner in which they did so deviated from the method prescribed by the scenario approach (Van Koppen & Mackor, 2019). Although the judges clearly constructed alternative scenarios, these did not all seem to be based on indications from the case file. Based on the written decision, it seems that the main aim was to confirm the scenario of guilt of the defendant, while trying to falsify the alternative scenario. Therefore, the legal practice deviates from the process of inference to the best explanation (IBE; Harman, 1965; Van Koppen & Mackor, 2019). According to IBE, the scenario that gives the best explanation for the evidence is considered to be most likely. The court seems not to have compared the scenarios to determine which scenario offered the best explanation for the evidence. Upon studying the case file, it became clear that a lot of evidence that supported scenarios of alternative perpetrators had not been included in the written decision. Similarly, evidence that contradicted the scenario of guilt of

the defendant had also not been included. When taking into account that the written judgement is not necessarily a reflection of the evidence considered by judges (Reijntjes & Reijntjes-Wendenburg, 2018), the possibility that the evidence against the scenario was considered but not included in the written decision, cannot be ruled out.

The case study therefore demonstrated that the use of falsification in legal practice, and the consideration of alternative scenarios, deviated from theory on those principles. Moreover, the lack of attention to exonerating evidence in the written decisions can be interpreted as an indication of confirmation bias, whereby the court's integration and interpretation of the evidence was tainted by their belief in the suspect's guilt. Hence, based on the case study, it seems the decision-making process may be at risk of insufficiently addressing factors with a negative influence on the decision made, despite attempts to protect against confirmation bias. These findings can also be related to, for instance, the Schiedammer Park murder, an established miscarriage of justice in the Netherlands in which tunnel vision was found to be a contributing factor (Posthumus, 2005). Based on the findings of the case analysis, exonerating evidence or alternative scenarios were insufficiently considered by the judges.

7.2.1.2. Survey. The issues that came to light through the case study were given more nuance in the second study, Chapter 3, using a survey of judges. Despite the small sample size, the results of the study identified a number of key areas where judges' experience and approach in practice differed from what would be required according to the theoretical background, such as the need for attempted falsification (Popper, 1959/2005) and the scenario approach (Van Koppen & Mackor, 2019). These include the consideration of alternative scenarios, where judges expressed a tendency to exclude alternative scenarios. The general consensus again indicated that evidence against alternative scenarios should be sought by the decision-maker rather than that the scenario which best explains the evidence should be deemed most likely. Sauerland and colleagues (2020) found that looking for evidence for and against alternative scenarios does

not protect against bias in the same way as identifying the scenario that best explains the evidence is thought to (Sauerland et al., 2020). Therefore, the way in which scenarios are considered and compared appears to affect the impact of using alternative scenarios at all. The scenario with the most support based on the evidence is not necessarily the scenario that offers the best explanation for the evidence, as there may also be evidence in the case file that contradicts that scenario. Such contradicting evidence would not reduce the evidence supporting the scenario, but does limit the extent to which the scenario can explain the evidence.

Respondents also showed a wide range in the extent to which they felt the need to take on an active role in trial proceedings. Although it is commonly accepted that the judge plays an active role, there is no clear definition of how that role should be fulfilled (Ferdinandusse, 2018). One promising finding from the survey was that respondents performed relatively well on the Wason card selection task and other tasks testing their application of falsification, compared to other groups (Rachlinski, 2012). However, when asked whether they felt they are able to sufficiently attempt falsification in case files in practice, the majority of the respondents felt they could not do so. Therefore, although the majority of judges seemed to understand the need for falsification, its application in practice does not currently appear to be in line with the best practice according to the theoretical understanding. In addition to that the consideration of alternative scenarios is mainly focused on excluding other possible explanations contrary to the indictment, rather than finding the best explanation (Van Koppen & Mackor, 2019), respondents also indicated that not every case file gave them the opportunity to use falsification, or to consider evidence contrary to the indictment.

7.2.1.3. Implications from the Research on Legal Practices. Based on the case study and the survey, it is clear that the majority of judges had a relatively good understanding of falsification and the need to consider alternative scenarios. However, judges themselves admitted to not feeling able to use falsification in

practice. From the case study, it also appeared that exonerating evidence in the case file was not taken into consideration, even when alternative scenarios were described. Although judges showed a range in the extent to which they felt the need to use falsification, related to their interpretation of an active judge, there appears to be a struggle for judges to use falsification in their decision-making process in practice. The current thesis contains insufficient information to allow me to clearly identify the causes of that struggle, but I emphasise that further research is needed. Some things that were mentioned by judges included a lack of time and the construction of the case file. As part of the experimental research reported in this thesis, the latter factor was studied.

7.2.2. Facilitating Falsification: Experimental Studies

In the other three studies that were conducted for the current thesis, I explored whether changes to the legal proceedings could facilitate falsification. The three studies focused on: changing the order in which the evidence is presented to the judge, changing instructions for the reasoned decision required from the judge, and using the Analysis of Competing Hypotheses (Heuer, 1999) to process the evidence in a case. However, studying whether confirmation bias could be countered proved to be difficult, as participants did not display unequivocal signs of confirmation bias.

7.2.2.1. Order Effects. In the study on order effects, the initial impressions of guilt were in line with the evidence presented to the participants. I expected participants to adhere to that initial belief. However, participants changed their perception of guilt appropriately in response to additional evidence that was presented. Participants did not show a disproportionate preference for information in line with the first piece of evidence in any of the conditions. Participants perceived the evidence according to its intended valence, and there was no disproportionate influence of the first piece of evidence. Instead, the perception of guilt, measured by the ratings for the likelihood of the suspect being guilty, was disproportionally affected by the last piece of evidence. That

recency effect did, however, not transfer to the absolute decision on whether or not to convict the suspect. Therefore, the order in which evidence is presented was argued to affect the perception of evidence, but not the integration of evidence.

One unexpected finding of the order effects study was that participants rated the exonerating evidence as more important than the incriminating evidence, regardless of the condition. Based on that finding, one reason why the manipulation did not cause a preference in line with the initial perception of guilt may have been that participants were biased towards the suspect's innocence. I offered the explanation that the bias towards innocence may be due to the fact that the participants were law students, who may have a preference for exonerating evidence due to their potential future role as defence lawyers.

Despite the fact that the preference for exonerating information was contradictory to expectations, it seems encouraging that the first piece of evidence did not cause a bias towards either guilt or innocence. That finding is particularly relevant in relation to the fact that the case file presented to judges in the Netherlands contains mostly incriminating information (Crombag, 2017). However, the implications of the study are also limited, as the manipulations, although based on criminal law proceedings, could not mimic the stakes or the investment in the case that judges likely experience in real life. Subsequently, the cognitive dissonance experienced in response to conflicting evidence, which was relatively low in this study, is likely to be greater in real life, and may result in a greater need for consonance or increased bias (Festinger, 1957).

7.2.2.2. Decision Accountability. In this experimental study (Chapter 5), I investigated whether the instructions to have to explain a decision affect the consideration of evidence by participants. The study was inspired by extensive research on the effects of accountability on decision-making, in which ample support has been found for the positive effect of the requirement to explain a decision (e.g., Lerner & Tetlock, 2003). Although the requirement of judges to explain their decision serves several functions (Verbaan, 2016), the effect of

detailed instructions on *how* to explain a decision had not yet been studied in the context of legal decision-making. The instructions to explain the decision used in this study were based on the explanation requirements according to the Dutch Code of Criminal Procedure (the justification condition), the German Code of Criminal Procedure (the explication condition), and the principle of falsification. A fourth condition which gave as little instruction as possible was included as a control condition.

In this study, lay participants were not disproportionately influenced by the incriminating case file. The Dutch condition did not differ from the control condition in terms of the amount exonerating evidence they included in their written decision, while the explication and falsification control included significantly more exonerating evidence than the control condition. Although the amount of exonerating evidence included in the written decision was a significant positive predictor of the decision to acquit the suspect, there were no differences in conviction rates between the four conditions. Therefore, there may have been a discrepancy between what participants included in their written judgements and what participants considered in their decision-making. That finding would also be in line with the interpretation of the Dutch Code of Criminal Procedure, according to which the written decision needs to indicate evidence that the decision could reasonably rest upon, but does not have to be a reflection of the decision-making process (Reijntjes & Reijntjes-Wendenburg, 2018). Furthermore, the discrepancy between what was considered and what was included in the written decision could result in the written decision not serving the several functions it has (Verbaan, 2016).

7.2.2.3. Analysis of Competing Hypotheses. In the study on the Analysis of Competing Hypotheses technique (ACH; Heuer, 1999), neither condition showed a preference for information confirming the suspect's guilt, therefore making it impossible to determine whether or not the ACH method could protect against the bias. However, based on exploratory analyses, it became clear that few participants in the ACH condition actually created a matrix as instructed by the

method. Furthermore, the average helpfulness rating for the matrix was low. The findings in this study supplemented research by Dhami et al., who found that the ACH method may not be suitable as a method to reduce confirmation bias due to problems with its implementation (Dhami et al., 2019). In my study using ACH, participants also seemed to prefer the evidence that was exonerating. That was visible in both their importance rating of the evidence and their choice of further questions to ask, where they favoured the evidence that was exonerating for the main suspect. That preference for exonerating information was also in line with the findings from the order effects study.

Despite using similar methods to other studies which have successfully induced confirmation bias (e.g., Ask & Granhag, 2005; Rassin et al., 2010; Rassin, 2018), as well as extensive pretesting and participants' initial impression of the suspect as being guilty, participants did not show a preference for incriminating information in this study. One explanation for the lack of bias observed was thought to be due to the sample used in both these studies, namely law students. Although the current research could not adequately test whether ACH could protect against confirmation bias, it nevertheless supported the existing literature on the problems with implementing the ACH method.

7.2.2.4. Implications from the Experimental Research. The experimental research presented in the current thesis was aimed at determining ways in which falsification could be facilitated in the context of criminal law proceedings. That was tested by determining whether three different methods could increase the consideration of exonerating evidence, and so counter the presence of confirmation bias. Although participants in both the study on order effects and the study on the Analysis of Competing Hypotheses overall showed a preference for the exonerating evidence, that was not influenced by the experimental manipulations. The preference for exonerating information was unexpected, and seemed contradictory to findings from previous research (e.g., Rassin, 2018; Marksteiner et al., 2011). Although there appeared to be no effect of different instructions to explain the decision on the final decision on guilt, none of the

conditions appeared to be disproportionately influenced by the incriminating case file. Furthermore, the analysis of competing hypotheses, based on the study conducted here as well as other research, seems to require further development before it can be put to use in contexts other than intelligence analysis.

7.3. Overall Implications

Based on the combined findings from the research in practice and the experimental research, it appears that there is a discrepancy between the way in which evidence is considered by judges in practice, and the way in which it was considered by participants (mainly law students) in the experimental studies, as a preference for exonerating information was observed in several studies. In the order effects study, the fact that the initial case was neutral, that is, not excessively incriminating or exonerating, may have influenced the favourable perception of exonerating information. However, in the ACH study, the case was clearly incriminating, which was evident from both the pretests conducted and the initial ratings of guilt given by participants. Besides the lack of bias towards an initial belief, participants in both the order effects and the ACH study actually showed a preference for exonerating information. In the order effects study, that was clear from the importance ratings, while in the ACH condition, it was clear from the importance ratings as well as the questions selected by participants. In previous studies, inducing a belief in guilt, or having participants express a hypothesis of guilt, did create a bias towards incriminating information. For instance, O'Brien (2009) found that participants who formulated a hypothesis about a prime suspect were biased towards guilt when reading the rest of the case file. Marksteiner et al. (2011) found that those who received an incriminating version of a case file favoured incriminating evidence compared to those who received an innocence version of the case. Therefore, the materials which were used, that were incriminating as determined by the pretests, does not seem to offer a sufficient explanation for the preference for exonerating information that was observed.

Another explanation for the favourable perception of exonerating evidence compared to incriminating evidence was the use of law students as participants. Law students were used due to their affinity with the context of the study, and the fact that judges were practically impossible to recruit as participants. However, law students are generally also trained for becoming defence lawyers, and may therefore have a keener interest for exonerating evidence than lay people. That explanation has unfortunately, to my knowledge, not been tested empirically. However, the deviant response by law students has also been observed in other experimental studies (e.g., Rassin et al., 2010). Therefore, there is relative support for the explanation that there was no preference in line with an initial belief due to the sample used for these experimental studies.

There appears to be a difference between law students and professional judges in their approach to evidence. There are several possible explanations for that difference. One is that the consideration of evidence as practiced by law students fades as judges gain experience in practice. Researchers in other areas have found the impact of training to diminish over time (Clarke et al., 2011). The same may also be true for those trained as judges, although, based on the interviews and the survey, it seems the need to engage in falsification and consider alternative scenarios is an individual need rather than one that has been installed in judges by training. Furthermore, judges also undergo additional years of training compared to law students. The training itself, in combination with the selection procedure for the training, may result in a limited variance among judges. Therefore, an effect that is observed in law students might be less likely to be observed in judges, due to the less diverse sample.

Another potential explanation for the difference in consideration of exonerating evidence between professional judges and participants in the current studies could be the difference in how the evidence was presented. The material in the experimental research I conducted was designed in such a way that, although at times biased towards the guilt of the suspect, the exonerating evidence was still easily accessible to participants. In cases presented to judges,

the information that could potentially be exonerating is likely not included in the mainly incriminating file, or in the summary of the file (Crombag, 2017). The fact that information to test alternative scenarios was often lacking in the case file was also mentioned by respondents in the survey. That difference between case files and the material used in the experimental studies may also have caused the lack of difference between experimental groups in several of the experimental studies. The obvious suggestion of an alternative scenario may have made all participants aware of the need to consider the alternative scenario, thereby reducing the potential impact of the manipulations. Although that could have resulted in participants giving desirable answers, it could nevertheless support the value of including more exonerating evidence, or more investigation of alternative scenarios, in the case file presented to the judge. Similarly, enabling the defence to conduct further investigation may also contribute to the consideration of the exonerating evidence by judges. Based on the responses to the survey, it appears judges would also welcome such a change. It should be noted that this finding is particularly relevant to the Dutch legal system, and the inquisitorial system. The conclusions drawn here may not apply to all jurisdictions, and may not be applicable to adversarial systems in which the role of the judge differs from the role of the judge in the inquisitorial system (Strier, 1992; Van Koppen & Penrod, 2003).

When reasoning in tasks that are routine, and when there is time constraint, the accuracy of decision-making can be adversely affected by the use of intuitive cognitive operations. When feedback is not provided, experience can contribute to faulty thinking in experts (Kahnemann, 2011; Tay et al., 2016). Choudhry et al. (2005) conducted a review of physicians' experience and the quality of the care they provide. According to their findings, those who have been in practice longer are likely to give care of a lower quality, and that interventions may be needed in order to improve or maintain the quality of care. When translating these findings to the tasks of judges, they seem to be at a similar risk as physicians. Judges are rarely provided with critical feedback on their decision-making

beyond colleagues coming to a different decision, for example, in appeal. In addition, many judges have several years of experience of deciding on cases, some of which may be very similar. Judges likely also have the experience that the majority of defendants who are brought to trial are guilty (Crombag, 2017). Therefore, the quality of their decision-making may be at risk.

Furthermore, the differences observed could also be due to the stress that the judicial system in the Netherlands is currently experiencing. Concerns have been raised on several occasions by Tegenlicht (2018) and the Dutch Association for the Administration of Justice (2018) relating to a lack of time and resources affecting the quality of the judicial process. Judges are expected to decide on a large amount of cases in a relatively short amount of time, and 40% of judges were found to work more than they should according to the conditions of their employment (Capgemini, 2019). It is generally accepted that a lack of time can result in so-called heuristic thinking (Kahnemann, 2011), which is more vulnerable to bias. Although the effect of such heuristic thinking may be economical, it can also be a contributing factor to errors (Tversky & Kahnemann, 1974), such as a miscarriage of justice when a defendant is innocent. Engel and Gigerenzer (2006) have also argued that heuristics are needed in complex or uncertain situations, as is often the case for legal fact finders. They further acknowledge that heuristics also ignore information, which is why the preferred strategy would capture the right information and ignore the rest. Hence, it would be extremely valuable to find a way in which the efficiency of legal decision-making can be preserved without causing detrimental biased thinking. It would be recommended to limit the use of heuristic thinking, for instance, by developing and implementing training to promote analytic thinking. Similar training has previously been researched by Morewedge and colleagues (2015) and Shaw and colleagues (2018), although its implementation and use in specific contexts has not yet been investigated. Another recommendation would be to resolve the time deficiency that likely results in biased thinking, which could, for example, be achieved through increased resources.

7.4. Limitations and Generalisability

A number of limitations can be identified in the research conducted for the current thesis. In the first two studies, where I examined falsification in practice, the limitation was largely due to the limited sampling of case materials and participants. Of course, the case study only focused on one case, and the insight it could give into legal decision-making in general was therefore limited. Nevertheless, studying a single case allows for useful insights that might otherwise be difficult to achieve as it can offer an in-depth evaluation of issues that is unlikely to be gained through other sources. The case at hand was a case that was brought to the Project Reasonable Doubt, and therefore, it had to meet certain requirements. For instance, the convicted individual must claim that he is innocent, because the crime was committed by someone else or because no crime took place (Amsterdam Lab for Legal Psychology, n.d.). The case also has to allow for a team of students to conduct an analysis within a given time frame. The case is therefore not representative of all cases that are decided by judges. For instance, a case in which a suspect provides a confession with valid perpetrator knowledge, would likely result in a different outcome than the one studied here. The chosen case was well suited for the aims of the current research as it clearly outlined the several scenarios considered by the court. Therefore, the reasoning could be compared to the alternative scenario approach (Van Koppen & Mackor, 2019). In order to determine how representative the current case study was of other decisions, future research could include other cases, both for other types of crimes and for decisions with a different outcome.

The sample used for the survey also had some limitations. Despite distributing the survey through several channels, there was a very low response rate to the survey. The judges who were initially approached to take part had, in Van Veldhuizen et al. (2019), agreed to being contacted to take part in future research. It may therefore be possible that the judges who answered the survey had more interest in research than the average judge. Therefore, their answers may not be generalisable to all criminal law judges in the Netherlands. Furthermore,

although the survey identified that there are issues for judges in applying falsification in practice, it did not sufficiently clarify what those issues are and how they could be resolved. Future research should try and further elucidate the issues which prevent the use of falsification in practice.

There are several limitations that are applicable to all experimental studies. The main limitation is the fact that I could not use real judges as participants. Of course, the training and experience of real judges differs from that of law students. The influence of expertise on confirmatory processing has previously been investigated by Schmittat and Englich (2016). They found that, in criminal law cases, criminal law experts showed less bias in their thinking than did other legal experts. Therefore, there appears to be a difference caused by the more specific training in the area of criminal law. However, based on the finding by Schmittat and Englich (2016), law students without a specific expertise might be expected to be more biased, which was not observed in the findings from the current research.

In addition to the failure to include real judges, the second limitation was the lack of ecological validity for the case file used and the importance of the decision. Although the use of case files similar to the ones used in real trials would likely present other difficulties due to the effort it would require from participants, the research may have benefitted from improved ecological validity. For instance, the method used by Schünemann and Bandilla (1989), whereby real judges were presented with case files based on real cases, could greatly benefit future research into the reasoning process used by judges. The use of a mock case with low stakes could, in the current research, have made it difficult for participants to commit to their initial impression of guilt, or for them to be invested in their decision. That could also be another explanation as to why participants, contrary to expectations, did not seem to be influenced by their initial impressions of guilt.

The use of a single case could also be considered a limitation of the current research. Although the case I used was pretested extensively before each of

the studies, findings based on a single case cannot be conclusively generalised. However, even with a single case, psychological mechanisms may be identified which are likely to be applicable to other cases as well. The type of case should not impact the decision-making process, as the minimum amount of evidence legally required remains the same, and the ease with which a judge becomes convinced should not differ depending on the crime. It may, however, be possible that other factors, such as emotional involvement, influence the cognitive processes used by judges in their decision-making of specific cases. The generalisability of findings should be determined by using other stimulus cases. In order to detect whether the type of case makes a difference, significantly more participants would be required. Hence, for the scope of this thesis, it was considered most efficient to only use one case. Although not necessarily a limitation, it can be argued that the focus of this thesis is restricted to approaches to legal decision-making that can be categorised as story-based approaches. However, there are plenty of other models of legal decision-making which warrant further research. A different approach worth discussing due to its recently gained popularity is the use of Bayesian thinking in legal decision-making. According to Bayesian models of thinking, the decision maker has to reassess the probability of a hypothesis in response to new evidence, using the statistical likelihoods of the evidence and the hypotheses (Dahlmann, 2019). However, there is resistance from legal professional to using the Bayesian approach, due to the fact that it is often subjective (e.g., the determination and updating of likelihoods) when used in a legal case (Fenton et al., 2019). Furthermore, even proponents of Bayesian methods acknowledge that the approach requires considerable effort to learn, which may be difficult for legal professionals due to its fairly complicated use of mathematics (Dahlmann, 2019). In a study on the use of Bayesian reasoning in juror decision-making, mock jurors displayed a reluctance to base causal inferences on statistical information was observed (Faigman & Baglioni, 1999). Therefore, although the use of Bayesian thinking may lead to pure logical

reasoning, it is far removed from the type of reasoning in which judges actually engage (Roberts, 2019), and was therefore not considered in the current thesis.

Lastly, as previously mentioned, the generalisability of the research conducted for this thesis may be limited by the focus placed on the Dutch or inquisitorial system. The case study and survey studies were both based in the Netherlands, and so the findings from these studies should not be extended to other jurisdictions without additional research. Similarly, the study on accountability was also based on inquisitorial systems. There are several differences between adversarial systems and inquisitorial systems which affect the factors that were studied in the current thesis. The main difference which affects the generalisability of the current findings to other systems is that the role of the judge differs, as the judge in the adversarial system does not decide or investigate in the same way as the judge in the inquisitorial system does (Strier, 1992; Van Koppen, 2007; Spencer, 2016). As the research conducted for the current thesis related specifically to the decision-making by judges in the inquisitorial system, the interpretation of the findings should also be limited to that context and interpretations with regards to other context should take into account the differences between the systems.

7.5. Areas for Future Research

One area for future research that became apparent during the research conducted for the current thesis is the discrepancy between the results observed in law students in two experimental studies, as well as some of the simple tasks done by judges in the survey study, and the application in practice. Based on the case study, it seems that the consideration of alternative scenarios in practice is not done in such a way that it can protect against confirmation bias. A similar impression emerged from the open answers given by respondents in the survey. The question then becomes what issues prevent judges in practice from the thorough consideration of alternative scenarios, or from conducting investigations into the exonerating evidence in a case file.

In line with that question, the survey also revealed another area for future research, namely the concept of an active judge. Although an active judge is expected in the inquisitorial system (Strier, 1992), there is no commonly accepted interpretation among judges of what an active judge is required to do. How active a judge is can for instance relate to the ordering of further investigative measures, or the consideration of alternative scenarios that were not proposed by the defence. As there is a lot of variation in how the role of an active judge is fulfilled in practice (Ferdinandusse, 2018; De Weerd, 2013), it would be helpful to gain insight into how these variations can affect the consideration of evidence, or even the outcome of cases. One way in which that could be researched is through case file analyses, for instance determining whether there is a relationship between the additional investigations ordered and the inclusion of alternative scenarios in the final decision. Another valuable source would be to interview a number of judges to determine how they fulfil the role of an active judge, and to look at whether that is related to their approach to additional investigations or consideration of the evidence leading to the final decision.

Based on the current research, it is hypothesised that law students have a preference for exonerating information as they have recently been trained to reason not only from the standpoint of the judge, but also from that of the defence lawyer. Therefore, when thinking about the suspect's guilt, they may inherently be focused on finding evidence against the suspect's guilt. In order to determine whether that explanation can account for the findings, future research should determine whether defence lawyers do have a preference for exonerating evidence compared to judges. Although the explanation seems logical when considering their tasks, the hypothesis should be tested empirically. However, that research may encounter a similar obstacle as the current research, namely difficulties in recruiting a sufficient number of research participants.

Another factor of legal decision-making that can be studied in relation to the consideration of exonerating evidence is the role of emotions. Although it was not accounted for in the current research, researchers have previously

identified several ways in which emotions can impact legal decision-making. For instance, anger has been associated with greater susceptibility to heuristic cues (Tiedens & Linton, 2001) and with offensive behaviour being perceived as more intentional (Ask & Pina, 2011). Several researchers have also considered the role of emotion in blame attribution (Alicke, 2000; Feigenson & Park, 2006). In future research, it could therefore be explored how emotion interacts with factors studied in the current research, such as for instance whether the extent to which alternative scenarios are considered is affected by the active decision-maker's emotional state.

7.6. Conclusion

The current thesis aimed to elucidate whether and how falsification is used by judges in their decision-making, and to examine ways in which falsification could be facilitated. The research into judges' practice found that judges are clearly trying to protect against the natural tendency of confirmation bias, mainly by the consideration of alternative scenarios. However, it also became clear that that is often not done in such a way that it is efficient in protecting against confirmation bias, and that judges struggle to use falsification in practice. As in several of the studies, participants' evaluation and selection of evidence did not seem to be influenced by their initial perception of guilt, it could not be determined if the experimental manipulations could protect against bias. Participants unexpectedly favoured exonerating evidence in both the study on order effects and the study on the ACH method, which may have been due to using law students as participants, although that explanation remains to be tested. The current research identified several areas for further research in practice on judges' decision-making. That field of research should be developed further in order to increase the effective use of falsification and the consideration of alternative scenarios in criminal law proceedings.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53*(1), 63–70. https://doi.org/10.1016/j.jom.2017.06.001

Adaval, R., & Wyer Jr, R. S. (1998). The role of narratives in consumer information processing. *Journal of Consumer Psychology*, *7*(3), 207-245. https://doi.org/10.1207/s15327663jcp0703_01

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*(4), 556–574. https://doi.org/10.1037/0033-2909.126.4.556

Amsterdam Lab for Legal Psychology (n.d.). *Reasonable Doubt.* https://allp.nl/projectgeredetwijfel/

Anderson, T., Schum, D., & Twining, W. (2005). *Analysis of evidence*. Cambridge University Press.

Ask, K., & Granhag, P. A. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, *2*(1), 43-63. https://doi.org/10.1002/jip.19

Ask, K., & Pina, A. (2011). On being angry and punitive: How anger alters perception of criminal intent. *Social Psychological and Personality Science, 2*(5), 494–499. https://doi.org/10.1177/1948550611398415

Ask, K., Rebelius, A., & Granhag, P. A. (2008). The 'elasticity' of criminal evidence: a moderator of investigator bias. *Applied Cognitive Psychology*, *22*(9), 1245–1259. https://doi.org/10.1002/acp.1432

Ask, K., Reinhard, M. A., Marksteiner, T., & Granhag, P. A. (2011). Elasticity in evaluations of criminal evidence: Exploring the role of cognitive dissonance. *Legal and Criminological Psychology, 16*(2), 289-306. https://doi.org/10.1348/135532510X510153

Bandes, S. (2006). Loyalty to one's convictions: The prosecutor and tunnel vision. *Howard Law Journal, 49*(2), 475-494.

Baron, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge University Press.

Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes*, *42*(1), 88-110. https://doi.org/10.1016/0749-5978(88)90021-0

Bennett. W. L. & Feldman, M. S. (2014). *Reconstructing reality in the courtroom, justice and judgement* (2nd ed.). Quid Pro Books. (Original work published 1981)

Bex, F. J. (2010). *Evidence for a good story: A hybrid theory of arguments, stories and criminal evidence*. [Doctoral dissertation, University of Groningen]. Springer.

Bex, F. J., Van Koppen, P. J., Prakken, H., & Verheij, B. (2010). A hybrid formal theory of arguments, stories and criminal evidence. *Artificial Intelligence and Law, 18*(2), 123-153. https://doi.org/10.1007/s10506-010-9092-x

Burke, A. (2007). Neutralizing cognitive bias: An invitation to prosecutors. *New York University Journal of Law Liberty, 2*, 512-530.

Buruma, Y. (2009). Betrouwbaar bewijs.[Reliable evidence]. *Delikt en Delinkwent, 4*, 303-324.

Cancino Montecinos, S. (2020). *New perspectives on cognitive dissonance theory* [Doctoral dissertation, University of Stockholm]. http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-179559

Capgemini (2019). *Het tijdsbestedingsonderzoek in relatie tot de productie-gerelateerde bijdrage voor het primair proces van de rechtspraak [The research on use of time in relation to production-related contribution to the primary process of the judiciary].* Rechtspraak. https://www.rechtspraak.nl/SiteCollectionDocuments/2019%20tijdbestedingsonderzoek%20eindrapport.pdf

Chang, W., Berdini, E., Mandel, D. R., & Tetlock, P. E. (2017). Restructuring structured analytic techniques in intelligence. *Intelligence and National Security*, *33*(3), 337–356. https://doi.org/10.1080/02684527.2017.1400230

Charman, S.D., Bradfield Douglass, A., & Mook, A. (2019). Cognitive bias in legal decision making. In N., Brewer, & A., Bradfield Douglass (Eds.) *Psychological Science and the Law,* (pp. 30-53). The Guilford Press.

Charman, S. D., Carbone, J., Kekessie, S., & Villalba, D. K. (2016). Evidence evaluation and evidence integration in legal decision-making: Order of evidence presentation as a moderator of context effects. *Applied Cognitive Psychology*, *30*(2), 214-225. https://doi.org/10.1002/acp.3181

Choudhry, N. K., Fletcher, R. H., & Soumerai, S. B. (2005). Systematic review: The relationship between clinical experience and quality of health care. *Annals of Internal medicine*, *142*(4), 260-273. https://doi.org/10.7326/0003-4819-142-4-200502150-00008

Clarke, C., Milne, R., & Bull, R. (2011). Interviewing suspects of crime: The impact of PEACE training, supervision and the presence of a legal advisor. *Journal of Investigative Psychology and Offender Profiling*, *8*(2), 149–162. https://doi.org/10.1002/jip.144

Cleiren, C.P.M. (2008). Het monopolie op de strafrechtelijke waarheidsvinding [The monopoly on truth-finding in criminal law]. *Strafblad, 6*(3*),* 272-285.

Cleiren, C. P. M. & Dubelaar, M. J. (2014). De betekenis van het scenario-denken voor het bewijs op grondslag van de tenlastelegging en de rechterlijke onderzoeksplicht [The meaning of scenario-thinking for the evidence based on the indictment and the judicial duty to investigate]. *Strafblad, 12*(6), 445-446.

Coen Verbraak (Host & Director), Hobbelink, M. (Producer) & Bollebakker, W. (Producer) (2015, August 15). Twijfels? (Season 8, Episode 4) [TV series episode]. In Francissen, W. (Executive Producer), *Kijken in de Ziel*. Skyhigh TV; NTR.

Cohen, M. (2015). When judges have reasons not to give reasons: Comparative law approach. *Washington and Lee Law Review, 72*(2), 483-572.

Cole, S. A. (2009). Cultural consequences of miscarriages of justice. *Behavioral Sciences & the Law*, *27*(3), 431–449. https://doi.org/10.1002/bsl.874

Convertino, G., Billman, D., Pirolli, P., Massar, J. P., & Shrager, J. (2008). The CACHE study: Group effects in computer-supported collaborative analysis. *Computer Supported Cooperative Work (CSCW)*, *17*(4), 353–393. https://doi.org/10.1007/s10606-008-9080-9

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*(3), 187–276. https://doi.org/10.1016/0010-0277(89)90023-1

Costabile, K. A., & Klein, S. B. (2005). Finishing strong: Recency effects in juror judgments. *Basic and Applied Social Psychology, 27*(1), 47-58. https://doi.org/10.1207/s15324834basp2701_5

Crombag, H.F.M. (2017). Invloeden op rechterlijke beslissingen [Influences on judicial decisions]. In P.J. van Koppen, J.W. de Keijser, R. Horselenberg, & M. Jelicic (Eds.) *Routes van het recht [Routes of the law]*, (pp. 385-397). Boom Juridisch.

Crombag, H. F. M., & Wagenaar, W.A. (2000). Audite et alteram partem. *Trema, 23*(3), 93-96.

Crombag, H.F.M., van Koppen, P.J., & Wagenaar, W.A. (2006). *Dubieuze zaken: De psychologie van strafrechtelijk bewijs. [Dubious cases: The psychology of evidence in criminal law]*. Olympus. (Original work published 1992).

Dahlman, C. (2019). De-biasing legal fact-finders with Bayesian thinking. *Topics in cognitive science*, advanced online publication. https://doi.org/10.1111/tops.12419

De Keijser, J.W. (2017). *Als de waarheid eraan moet geloven: Alledaagse bedreigingen voor waarheidsvinding in het strafproces [When the truth is taken down: Everyday threats for truth-finding in criminal law procedure].* Boom Juridisch.

De Langhe, B., van Osselaer, S. M. J., & Wierenga, B. (2011). The effects of process and outcome accountability on judgment process and performance. *Organizational Behavior and Human Decision Processes*, *115*(2), 238–252. https://doi.org/10.1016/j.obhdp.2011.02.003

De Weerd, L. (2013). Geen scheidsrechter maar rechter [Not referee but judge], *Trema, 36*(5), 156-160.

Dhami, M. K., Belton, I. K., & Mandel, D. R. (2019). The "analysis of competing hypotheses" in intelligence analysis. *Applied Cognitive Psychology*, *33*(6), 1080–1090. https://doi.org/10.1002/acp.3550

Dreissen, W.H.B. (2007). *Bewijsmotivering in strafzaken [Reasoned decisions in criminal law cases].* Boom Juridisch.

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*(1), 74-78. https://doi.org/10.1016/j.forsciint.2005.10.017

Dubelaar, M.J. (2014). *Betrouwbaar getuigenbewijs: waardering en totstandkoming van strafrechtelijke getuigenverklaringen in perspectief [Reliable witness evidence: evaluation and formation of criminal law witness statements in perspective]* [Doctoral dissertation, Leiden University] http://hdl.handle.net/1887/23625

Eerland, A., & Rassin, E. (2012). Biased evaluation of incriminating and exonerating (non)evidence. *Psychology, Crime & Law*, *18*(4), 351–358. https://doi.org/10.1080/1068316x.2010.493889

Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology, 67*(3), 382-394. https://doi.org/10.1037/0022-3514.67.3.382

Engel, C., & Gigerenzer, G. (2006). Law and heuristics: An interdisciplinary venture. In G.

Gigerenzer & C. Engel (Eds.), *Heuristics and the law* (pp. 1-16). MIT Press.

European Court of Human Rights (2019). *Guide on Article 6 of the European Convention on Human Rights: Right to a fair trial (criminal limb).*

Faigman, D. L., & Baglioni, A. J. (1988). Bayes' theorem in the trial process: Instructing jurors on the value of statistical evidence. *Law and Human Behavior, 12*(1), 1–17. https://doi.org/10.1007/bf01064271

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Feigenson, N., & Park, J. (2006). Emotions and attributions of legal responsibility and blame: A research review. *Law and Human Behavior, 30*(2), 143-161. https://doi.org/10.1007/s10979-006-9026-z

Fenton, N., Neil, M., Yet, B., & Lagnado, D. (2019). Analyzing the Simonshaven case using Bayesian networks. *Topics in cognitive science,* advanced online publication. https://doi.org/10.1111/tops.12417

Ferdinandusse, W.N. (2018). De actieve rechter [The active judge], *Strafblad, 16*(2), 22-40.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Findley, K. A. & Scott, M. S. (2006). Multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review*, 2, 291-398.

Gommer, H. (2007). Onbewuste denkprocessen maken motivering tot noodzaak [Subconcious thought-processes necessitate reasoned decisions]. *Trema, 30*(4), 127-134.

Green, M. C., & Donahue, J. K. (2011). Persistence of belief change in the face of deception: The effect of factual stories revealed to be false. *Media Psychology, 14*(3), 312-331. https://doi.org/10.1080/15213269.2011.598050

Greenspan, R., & Scurich, N. (2016). The interdependence of perceived confession voluntariness and case evidence. *Law and Human Behavior, 40*(6), 650–659. https://doi.org/10.1037/lhb0000200

Hall, A. T., Frink, D. D., & Buckley, M. R. (2015). An accountability account: A review and synthesis of the theoretical and empirical research on felt accountability. *Journal of Organizational Behavior, 38*(2), 204–224. https://doi.org/10.1002/job.2052

Harman, G. (1965). The inference to the best explanation. *Philosophical Review, 74*(1), 88–95. https://doi.org/ 10.2307/2183532

Harmon-Jones, E., & Harmon-Jones, C. (2007). Cognitive dissonance theory after 50 years of development. *Zeitschrift für Sozialpsychologie, 38*(1), 7-16. https://doi.org/10.1024/0044-3514.38.1.7

Heuer, R. J. (1999). Analysis of competing hypotheses. In *Psychology of Intelligence Analysis* (pp.95 -110). Central Intelligence Agency - Center for the Study of Intelligence. https://doi.org/10. 1007/s10606-008-9080-9

Innocence Project (2020) *All Cases.* https://www.innocenceproject.org/all-cases/

Insko, C. A. (1964). Primacy versus recency in persuasion as a function of the timing of arguments and measures. *The Journal of Abnormal and Social Psychology, 69*(4), 381-391. https://doi.org/10.1037/h0042765

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology, 80*(4), 557-571. https://doi.org/10.103700022-3514.80.4.557

Jones, M., & Sugden, R. (2001). Positive confirmation bias in the acquisition of information. *Theory and Decision*, *50*(1), 59-99. https://doi.org/ 10.1023/A:1005296023424

Jonas, E., Traut-Mattausch, E., Frey, D., & Greenberg, J. (2008). The path or the goal? Decision vs. information focus in biased information seeking after preliminary decisions. *Journal of Experimental Social Psychology, 44*(4), 1180-1186. https://doi.org/10.1016/j.jesp.2008.02.009

Kahnemann, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux

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

Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior*, *27*(2), 187-203. https://doi.org/10.1023/A:1022599230598

Kerstholt, J. H., & Jackson, J. L. (1998). Judicial decision making: Order of evidence presentation and availability of background information. *Applied Cognitive Psychology, 12*(5), 445-454. https://doi.org/10.1002/(SICI)1099-0720(199810)12:5<445::AID-ACP518>3.0.CO;2-8

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. https://doi.org/10.1016/j.jarmac.2017.09.001

Lehner, P. E., Adelman, L., Cheikes, B. A., & Brown, M. J. (2008). Confirmation bias in complex analyses. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *38*(3), 584-592. https://doi.org/10.1109/TSMCA.2008.918634

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255–275. https://doi.org/10.1037/0033-2909.125.2.255

Lerner, J.S., & Tetlock, P.E. (2003). Bridging Individual, Interpersonal, and Institutional Approaches to Judgment and Decision Making: The Impact of Accountability on Cognitive Bias. In S. Schneider & J. Shanteau (Eds.), *Emerging Perspectives on Judgment and Decision Research* (pp. 431-457). Cambridge University Press. https://doi:10.1017/CBO9780511609978.015

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764-766. https://doi.org/10.1016/j.jesp.2013.03.013

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098-2109. https://doi.org/10.1037/0022-3514.37.11.2098

Marksteiner, T., Ask, K., Reinhard, M. A., & Granhag, P. A. (2011). Asymmetrical scepticism towards criminal evidence: The role of goal-and belief-consistency. *Applied Cognitive Psychology*, *25*(4), 541-547. https://doi.org/10.1002/acp.1719

Martin, D. L. (2002). Lessons about justice from the laboratory of wrongful convictions: Tunnel vision, the construction of guilt and informer evidence. *UMKC Law Review, 70*(4), 847-864.

McDermott, Y. (2015). Inferential reasoning and proof in international criminal trials: The potentials of Wigmorean analysis. *Journal of International Criminal Justice*, *13*(3), 507-533. https://doi.org/10.1093/jicj/mqv020

Mendel, R., Traut-Mattausch, E., Jonas, E., Leucht, S., Kane, J. M., Maino, K., Kissling, W., & Hamann, J. (2011). Confirmation bias: Why psychiatrists stick to wrong preliminary diagnoses. *Psychological Medicine*, *41*(12), 2651-2659. https://doi.org/10.1017/S0033291711000808

Mevis, P.A.M. (2019). Modernisering van het strafprocesrecht op z'n Duits [Modernisation of criminal procedural law in the German way]. *Delikt & Delinkwent, 7*,40–47.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129-140. https://doi.org/10.1177/2372732215600886

Murukannaiah, P. K., Kalia, A. K., Telangy, P. R., & Singh, M. P. (2015). Resolving goal conflicts via argumentation-based analysis of competing hypotheses. *2015 IEEE 23rd International Requirements Engineering Conference (RE)*, 156–165. https://doi.org/10.1109/re.2015.7320418

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation Bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*(1), 85–95. https://doi.org/10.1080/00335557743000053

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

Nieuwsuur (8th November 2018) Brandbrief rechters: wij vrezen voor de toekomst van de rechtspraak. Retrieved from: https://nos.nl/nieuwsuur/artikel/2258390-brandbrief-rechters-wij-vrezen-voor-de-toekomst-van-de-rechtspraak.html.

O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology, Public Policy, and Law*, *15*(4), 315–334. https://doi.org/10.1037/a0017881

Pardo, M. S. & Allen, R. J. (2008). Juridical proof and the best explanation. *Law and Philosophy, 27*, 223-268. https://doi.org/10.1007/s10982-007-9016-4

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of personality and social psychology, 51*(2), 242-258. https://doi.org/10.1037/0022-3514.51.2.242

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 521-533. https://doi.org/10.1037/0278-7393.14.3.521

Pennington, N. & Hastie, R. (1991). Cognitive theory of juror decision making: The story model. *Cardozo Law Review, 13*(2), 519-558.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of Personality and Social Psychology*, *62*(2), 189-206. https://doi.org/10.1037/0022-3514.62.2.189

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition*, *49*(1), 123-163. https://doi.org/10.1016/0010-0277(93)90038-W

Pennington, J., & Schlenker, B. R. (1999). Accountability for consequential decisions: Justifying ethical judgments to audiences. *Personality and Social Psychology Bulletin*, *25*(9), 1067–1081. https://doi.org/10.1177/01461672992512001

Popper, K. (2005). *The logic of scientific discovery*. Routledge. (Original work published 1959)

Posthumus, F. (2005). *Evaluatieonderzoek in de Schiedammer parkmoord*. Openbaar ministerie. https://web.archive.org/web/20051016074222/https://www.om.nl/ pg/documents/ rapport.pdf

Price, H. L., & Dahl, L. C. (2014). Order and strength matter for evaluation of alibi and eyewitness evidence. *Applied Cognitive Psychology*, *28*(2), 143-150. https://doi.org/ 10.1002/acp.2983

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369-381. https://doi.org/ 10.1177/0146167202286008

Puvathingal, B. J., & Hantula, D. A. (2012). Revisiting the psychology of intelligence analysis: From rational actors to adaptive thinkers. *American Psychologist*, *67*(3), 199–210. https://doi.org/10.1037/a0026640

Rachlinski, J.J. (2012). Judicial Psychology. *Rechtstreeks, 2,* 15-34.

Rassin, E. (2010). Blindness to alternative scenarios in evidence evaluation. *Journal of Investigative Psychology and Offender Profiling*, *7*(2), 153-163. https://doi.org/ 10.1002/ jip.116.

Rassin, E. (2018). Fundamental failure to think logically about scientific questions: An illustration of tunnel vision with the application of Wason's card selection test to criminal evidence. *Applied Cognitive Psychology*, *32*(4), 506-511. https://doi.org/ 10.1002/acp.3417

Rassin, E. (2018). Reducing tunnel vision with a pen-and-paper tool for the weighting of criminal evidence. *Journal of Investigative Psychology and Offender Profiling*, *15*(2), 227–233. https://doi.org/10.1002/jip.1504

Rassin, E., Eerland, A., & Kuijpers, I. (2010). Let's find the evidence: an analogue study of confirmation bias in criminal investigations. *Journal of Investigative Psychology and Offender Profiling*, *7*(3), 231–246. https://doi.org/10.1002/jip.126

Reijntjes, J.M. & Reijnjes-Wendenburg, C. (2018). De Bewijsconstructie [The evidence construction]. In *Handboek Strafzaken 34: Bewijs* [EPUB version]. Retrieved from: https://www.navigator.nl

Roberts, P. (2019). Scenarios, probability, and evidence scholarship, old and new. *Topics in Cognitive Science*. Advance online publication. https://doi.org/10.1111/tops.12479

Sauerland, M., Krix, A. C., Merckelbach, H. (2016). Identificaties door ooggetuigen: Waarom een rechtspsycholoog handig is [Identifications by eyewitnesses: Why a legal psychologist is useful], *Nederlands Juristenblad, 10*(30), 2170-2176.

Sauerland, M., Otgaar, H., Maegherman, E., & Sagana, A. (2020) Allegiance bias in statement reliability evaluations is not eliminated by two-sided instructions. [Manuscript accepted for publication], *Zeitschrift für Psychologie*.

Schmittat, S. M., & Englich, B. (2016). If you judge, investigate! Responsibility reduces confirmatory information processing in legal experts. *Psychology, Public Policy, and Law, 22*(4), 386-400. https://doi.org/10.1037/law0000097

Schünemann, B. & Bandilla, W. (1989). Perseverance in courtroom decisions. In H. Wegener, F. Lösel, & J. Haisch (Eds.) *Criminal behavior and the justice system: Psychological perspectives* (pp. 181 – 192). Springer.

Shaw, A., Kenski, K., Stromer-Galley, J., Mikeal Martey, R., Clegg, B. A., Lewis, J. E., Folkestad, J.E., Strzalkowski, T. (2018). Serious efforts at bias reduction: The effects of digital games and avatar customization on three cognitive biases. *Journal of Media Psychology: Theories, Methods, and Applications, 30*(1), 16-28. https://doi.org/10.1027/1864-1105/a000174

Shimizu, T., Matsumoto, K., & Tokuda, Y. (2013). Effects of the use of differential diagnosis checklist and general de-biasing checklist on diagnostic performance in comparison to intuitive diagnosis. *Medical teacher*, *35*(6), 1218-1229. https://doi.org/10.3109/0142159X.2012.742493

Simmelink, J. B. H. M. (2001). Bewijsrecht en bewijsmotivering [Evidence law and reasoned decisions]. In M.S. Groenhuijsen & G. Knigge (Ed.), *Het onderzoek ter zitting* (pp.397-453). Rijksuniversiteit Groningen.

Simon, D. (2004). A third view of the black box: Cognitive coherence in legal decision making. *The University of Chicago Law Review*, *71*(2), 511-586.

Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*, *86*(6), 814-837. https://doi.org/ 10.1037/0022-3514.86.6.814

Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, *51*(3), 416–446. https://doi.org/10.1016/0749-5978(92)90020-8

Sorocham, D. J. (2008). Wrongful convictions: Preventing miscarriages of justice some case studies. *Texas Tech Law Review, 41*(1), 93-116.

Spencer, J. R. (2016). Adversarial vs inquisitorial systems: is there still such a difference? *The International Journal of Human Rights*, *20*(5), 601–616. https://doi.org/10.1080/13642987.2016.1162408

Stevens, L. (2014). Bewijs waarderen: Hoe doen strafrechters dat? [Evaluating evidence: How is that done by criminal law judges?]. *Nederlands Juristenblad, 8*(40), 2842-2850.

Stevens, L. (2017). *Onterechte vrijspraken? Over hoe de rechter redeneert met een alternatief scenario van de verdachte.* Boom Juridisch.

Stout, S. C., Amundson, J. C., & Miller, R. R. (2005). Trial order and retention interval in human predictive judgment. *Memory & Cognition, 33*(8), 1368-1376. https://doi.org/10.3758/BF03193369

Strier, F. (1992). What can the American adversary system learn from an inquisitional

system of justice. *Judicature, 76*(3), 109-111.

Tay, S. W., Ryan, P., & Ryan, C. A. (2016). Systems 1 and 2 thinking processes and cognitive reflection testing in medical students. *Canadian Medical Education Journal, 7*(2), 97-103.

Tenney, E. R., Cleary, H. M., & Spellman, B. A. (2009). Unpacking the doubt in "beyond a reasonable doubt": Plausible alternative stories increase not guilty verdicts. *Basic and Applied Social Psychology*, *31*(1), 1-8. https://doi.org/10.1080/01973530802659687

Tetlock, P. E. (1983). Accountability and the Perseverance of First Impressions. *Social Psychology Quarterly*, *46*(4), 282-292. https://doi.org/10.2307/3033716

Tetlock, P. E. (1985). Accountability: A Social Check on the Fundamental Attribution Error. *Social Psychology Quarterly*, *48*(3), 227–236. https://doi.org/10.2307/3033683

Tiedens, L. Z., & Linton, S. (2001). Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing. *Journal of Personality and Social Psychology, 81*(6), 973–988. https://doi.org/10.1037/0022-3514.81.6.973

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Twining, W. (1985). *Theories of evidence: Bentham & Wigmore*. Stanford University Press.

Van der Post, N., & Van Toor, D. (2019). Volgorde-effecten in het Nederlandse strafproces? Bayesiaanse methode als hulpmiddel tegen de invloed van het primacy en het recency effect op rechterlijke bewijswaardering. *Expertise en Recht, 11*(1), 41- 54.

Van Koppen, P.J. (2003). *De Schiedammer parkmoord: Een rechtspsychologische reconstructie*. Ars Aequi.

Van Koppen, P. J. (2007). Miscarriages of justice in inquisitorial and accusatorial legal systems. *Journal of the Institute of Justice and International Studies, 7,* 50-58.

Van Koppen, P. J. (2008). Blundering justice: The Schiedam Park Murder. In: R.N. Kocsis (red.), *Serial murder and the psychology of violent crimes* (pp. 207-228). Humana.

Van Koppen, P. J. (2011). *Gerede twijfel: Over bewijs in strafzaken*. De Kring.

Van Koppen, P. J. (2011). *Overtuigend bewijs: indammen van rechtelijke dwalingen*. Nieuw Amsterdam.

Van Koppen, P. J. & Penrod, S. D. (2003). Adversarial or inquisitorial: Comparing systems. In P.J. van Koppen & S.D. Penrod (Eds.), *Adversarial versus inquisitorial justice: Psychological perspectives on criminal justice systems* (pp. 2-20). Plenum.

Van Koppen, P.J. & de Keijser, W.J. (2017). Invloeden op rechterlijke beslissingen [Influences on judicial decisions]. In P.J. van Koppen, J.W. de Keijser, R. Horselenberg, & M. Jelicic (Eds.) *Routes van het recht [Routes of the law]*, (pp. 853-867). Boom Juridisch.

Van Koppen, P. J. & Mackor, A. R. (2019). A scenario-approach to the Simonshaven case. *Topics in Cognitive Science*, advanced online publication. https://doi.org/10.1111/tops.12429

Van Veldhuizen, T.S., Maegherman, E., & Horselenberg, R. (2019). Het verhaal gewogen: kennis van strafrechters over het geheugen en het politieverhoor [The story weighed: knowledge of criminal law judges about memory and the police interview]. *Strafblad, 17*(4), 58-63.

Van Westing, G. (14th June 2018). Unieke waarschuwing rechters en OM: 'Rechtspraak wordt uitgehold', 14th June 2018. Retrieved from: https://nos.nl/nieuwsuur/artikel/2250340-unieke-waarschuwing-rechters-en-om-rechtspraak-wordt-uitgehold.html

Verbaan, J.H.J (2016). *Straf(proces)recht begrepen [Criminal (procedural) law understood]*. Boom Juridisch.

Wagenaar, W.A., van Koppen, P.J. &, Crombag, H.F.M. (1993). *Anchored Narratives: The psychology of criminal evidence*. Harvester Wheatsheaf.

Wason, P. C. (1960). On the Failure to Eliminate Hypotheses in a Conceptual Task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129–140. https://doi.org/10.1080/17470216008416717

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*(3), 273–281. https://doi.org/10.1080/14640746808400161

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, *23*(1), 63-71. https://doi.org/10.1080/00335557143000068

Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning: Structure and content.* Harvard University Press.

Wells, G. L., & Lindsay, R. C. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88*(3), 776–784. https://doi.org/10.1037/0033-2909.88.3.776

Legal References

Wetboek van Strafvordering [Dutch Code of Criminal Procedure] artikel 339

Wetboek van strafvordering [Dutch Code of Criminal Procedure] artikel 359

Wetboek van strafvordering [Dutch Code of Criminal Procedure] artikel 360

Utrecht Court, 21st December 2012, ECLI:NL:RBUTR:2012:BY7099

Impact Paragraph

Throughout this thesis, the experimental research that has been conducted has been related to practice. Similarly, from the research on current practice, several conclusions were drawn about ways in which the current situation could potentially be improved. In this impact paragraph, I will emphasise the societal relevance of the research that was conducted, the audience for whom the findings are relevant, and further steps that can be undertaken to put the findings from this research to use in practice.

Miscarriages of justice have always caused widespread criticism, and have recently also become the topic of various popular culture programs (e.g., "Making A Murderer" or "When They See Us" on Netflix). When considering how a miscarriage of justice could have happened, there is often a focus on a particular piece of problematic evidence, such as a false confession or a wrongful identification. Accordingly, a lot of research has been conducted on how one piece of evidence can taint the rest of an investigation, as well as further legal proceedings. Related to the influence of incriminating, potentially problematic, evidence is also the lack of attention paid to evidence that could have been exonerating for the suspect. In many miscarriages of justice, there was also evidence that supported the suspect's innocence, such as an alibi, that was disregarded. It is of great value to society to understand how that evidence was then reasoned with to result in a wrongful conviction, and to identifying potential remedies to such wrongful decisions. The impact of a wrongful conviction affects not only the convicted individual and their family, but also the victim. Furthermore, the true perpetrator will not be punished for their crime, and wrongful convictions can also decrease the general trust in the criminal justice system. Despite the extensive impact on society, the research conducted here is most relevant to those working in the criminal justice system, and those who may attempt to reform it.

The consideration of exonerating evidence can be related to the concept of falsification, whereby evidence that disconfirms one's preferred theory or

hypothesis, should be sought. The research in this thesis was conducted in order to provide insight into the use of falsification by judges in legal decision-making in practice. The other main aim was to try and find ways in which falsification can be facilitated, which was done by conducting experimental studies based on criminal law proceedings using law students or members of the general public as participants. Based on two studies in practice, it became clear that judges often see the added value of looking for evidence that contradicts the charge, or to consider alternative scenarios, but that they struggle to do so in everyday cases. What exactly causes that struggle is an area for further investigation, although it became very clear that the role of an active judge is important in resolving that struggle. It is expected that a clarification of the role of an active judge would be beneficial for all parties involved in the legal arena. Related to the role of the active judge is also the role of the defence. As was mentioned several times by respondents in Chapter 3, the extent to which judges consider exonerating evidence or alternative scenarios is, for some, dependent on whether and how it is presented by the defence.

Although the experimental studies did not result in clear methods which can easily facilitate falsification, the findings did show a contrast between what was observed in practice, and the responses by participants in the experimental studies. Therefore, it would be useful to specifically identify the differences between law students and judges. For instance, further insight should be gained into what factors present obstacles for judges to use falsification in practice, and whether these are external factors related to the criminal justice system, or whether these are internal factors related to judges' personal experience at their task. The need for further insight is also related to the need for cooperation between practice and research. The lack of such a cooperation at the moment made it difficult for me to further develop my research and findings, and limited the conclusions that could be drawn from my research.

The research conducted for this thesis has been presented at various international conferences on the topic of legal psychology. A more extensive

version of the case study was also published in the Reasonable Doubt book series, which is available to, and written for, the general public. Furthermore, several of the chapters have also been submitted to peer-reviewed journals. The survey that was used in Chapter 3 has also been translated and distributed in Finland, and is currently also being translated to other languages for further dissemination. One of the subsequent aims is to incorporate findings from the international survey, along with the insight gained from this and other research, into an EU training protocol on safeguarding the presumption of innocence, which has been proposed to the European Commission.

Summary

The decision process that leads up to a judgement in a criminal law cases remains elusive. Although there are rules on the submission of evidence and the minimum amount of evidence required, very little is known about how evidence is weighed and integrated to reach a decision about whether or not the defendant committed the crime they are charged with. The charge essentially represents the theory that the defendant is guilty. One process that is considered a key step when trying to determine whether or not a theory is true is to look for evidence that disconfirms the initial theory, also known as falsification. Falsification can be closely related to the scenario approach, in which alternative scenarios are compared to find the scenario that best explains the evidence. However, the consideration of alternative scenarios can be hampered if the judge favours evidence that confirms the theory while paying disproportionately less attention to evidence that contradicts the theory. That tendency, known as confirmation bias, has repeatedly been shown to affect behaviour and decisions in various areas, including criminal investigations and legal decision-making. In the current thesis, the role of falsification in legal decision-making was investigated. After giving an overview of a selection of theories on legal decision-making in Chapter 1, two studies are reported, which tested the use of falsification in judges' practice (Chapters 2 and 3). In addition, three experimental studies were conducted in order to identify ways in which falsification could be facilitated (Chapters 4–6).

The study described in Chapter 2 is a case study of a murder case. The court's reasoned decision was analysed for the attempts at confirmation and falsification on the basis of the evidence in the case file. Based on the analysis, it appears evidence that supported the guilt of the defendant was taken into consideration disproportionally more so than evidence that supported the guilt of another suspect, or evidence that supported the defendant's innocence. However, the reasoned decision also contained several alternative scenarios that the court had considered. Therefore, it appears the court aimed to consider

alternative scenarios, but struggled to do so effectively and in accordance with the theoretical underpinnings of the scenario approach.

The second study was a survey conducted with criminal law judges, described in Chapter 3. The survey provided insight into judges' understanding of falsification, as well as their perceived need to apply falsification. Furthermore, respondents also described how they apply falsification in practice. Due to a low response rate, the answers for the survey were supplemented with information from interviews with criminal law judges. Overall, the findings from the survey were also in line with the findings from the case study, as participants showed a relatively good theoretical understanding of falsification, but experienced more difficulty when applying it. It also became clear that judges varied considerably in how they viewed their role, and the extent to which they felt the need to personally consider alternative scenarios or exonerating evidence.

In the study described in Chapter 4, the effects of the order of evidence presentation were studied. The initial piece of evidence was expected to influence the subsequent consideration of evidence. Contrary to expectations, the perceived likelihood of guilt appeared to be influenced mostly by the last piece of evidence presented to participants. However, that effect on guilt perception did not extend to the conviction rates. Overall, participants, who were law students, rated evidence that was exonerating for the suspect as significantly more important than the incriminating evidence, which was contrary to the expectations.

In the study reported in Chapter 5, the effect of different instructions on how to explain a decision on guilt was tested. In order to do so, four sets of instructions were created. These were based on the Dutch Code of Criminal Procedure, the German Code of Criminal Procedure, and the principle of falsification. Lastly, a control condition was included in which as little instruction was given as possible. Participants were given an incriminating case vignette. Due to the emphasis on the consideration of alternative explanations in the German Code of Criminal Procedure and in the principle of falsification, it was expected

that these conditions would have a lower conviction rate in comparison to the other conditions. However, there was no difference between conditions in their conviction rates. Nevertheless, participants in the German Code of Criminal Procedure condition and in the falsification condition included significantly more exonerating evidence in their written decision than did participants in the control condition. Based on the findings, the impact of different accountability instructions appears to affect the evidence that was included in the written decision, but not the final judgement on guilt. That could imply a discrepancy between the evidence that was considered by participants and the evidence that participants included in the written decision.

In the final experimental study (Chapter 6), it was investigated whether the Analysis of Competing Hypotheses — a technique for structured consideration of scenarios — could increase the consideration of exonerating evidence, and of alternative scenarios. Participants, who were again law students, received a case vignette, which described a clear main suspect, but also contained the suggestion of an alternative scenario. Contrary to expectations, participants in both the control and the ACH conditions favoured the exonerating information over the information that was incriminating for the main suspect. Furthermore, it appears that many participants did not apply the ACH properly, which supports findings from other researchers who have suggested that ACH may be too vaguely described for it to be applied properly.

The results of all studies are integrated and discussed in Chapter 7. The main conclusion on the basis of the findings is that judges are aware of the danger of confirmation bias, and the importance of falsification and the consideration of alternative scenarios. However, they seem to struggle to apply their knowledge in practice. One related implication is that there appears to be large variation in judges' perception of their active role, which could be a potential area where the use of falsification could be encouraged and improved. Lastly, although the use of law students may have limited the findings of two of the experimental studies, the observed preference for exonerating information suggests a discrepancy

between law students' decision making in experimental settings and judges' experience in practice. Another implication therefore may be that judges are limited in their use of falsification by restraints of the system that are outside of their control. That implication is also in line with the findings that judges appear to understand the theories on falsification and alternative scenarios, but struggle with applying the theories properly in practice.

Sammanfattning (Swedish Summary)

Beslutsprocessen bakom ett domslut i brottmål är svår att fånga. Även om det finns regler som styr bevisföringen och minimikrav på mängden bevis är mycket litet känt om hur bevis viktas och vägs samman för att nå ett beslut om huruvida den åtalade begått eller inte begått det brott den är anklagad för. Brottsanklagelsen utgår i grunden ifrån att den åtalade är skyldig. Ett särskilt viktigt steg i processen att försöka avgöra om en hypotes är sann eller inte är att leta efter bevis som motbevisar den ursprungliga hypotesen, även kallat falsifiering. Falsifiering är nära besläktat med scenarioansatsen, där olika scenarier jämförs för att finna det scenario som bäst förklarar bevismaterialet. Det kan emellertid vara svårare att ta hänsyn till alternativa scenarier om domaren ger företräde åt bevismaterial som bekräftar hypotesen men ägnar oproportionerligt lite uppmärksamhet åt bevismaterial som motsäger hypotesen. Denna tendens, kallad konfirmeringsbias, har upprepade gånger visat sig påverka beteende och beslut inom olika områden, inklusive brottsutredningar och rättsligt beslutsfattande. Föreliggande avhandling utforskade den roll falsifiering spelar inom rättsligt beslutsfattande. Efter en översikt över ett urval av teorier om rättsligt beslutsfattande i kapitel 1 redovisas två studier som testade användningen av falsifiering i domstolspraxis (kapitel 2 och 3). Dessutom genomfördes tre experimentella studier för att identifiera sätt att underlätta användande av falsifiering (kapitel 4–6).

Studien som beskrivs i kapitel 2 är en fallstudie av ett mordfall. Domstolens domskäl analyserades för att granska försök till konfirmering och falsifiering på grundval av bevismaterialet i handlingarna. Enligt analysen förefaller det som att bevismaterial som styrkte den åtalades skuld togs under övervägande i oproportionerligt hög grad, jämfört med bevismaterial som styrkte en annan misstänkts skuld, eller bevismaterial som styrkte att den åtalade var oskyldig. Emellertid innehöll domskälet också flera alternativa scenarier som domstolen hade tagit i beaktande. Det förefaller därför som att domstolen avsåg att ta

alternativa scenarier i beaktande men hade svårigheter att göra detta på ett effektivt sätt och i enlighet med de teoretiska grundvalarna för scenarioansatsen.

Den andra studien var en enkät som genomfördes med brottmålsdomare och som beskrivs i kapitel 3. Enkäten gav insikt om domares förståelse av falsifiering, liksom om deras upplevda behov av att tillämpa falsifiering. Vidare beskrev respondenterna hur de tillämpar falsifiering i praktiken. På grund av en låg svarsfrekvens kompletterades enkätsvaren med information från intervjuer med brottmålsdomare. Resultaten från enkäten överensstämde i huvudsak med resultaten från fallstudien, i det att deltagarna uppvisade relativt god teoretisk förståelse av falsifiering men upplevde tillämpningen därav som svårare. Det var också tydligt att domare skiljde sig väsentligt åt i fråga om hur de såg på sin roll och i vilken utsträckning de kände behov att personligen överväga alternativa scenarier eller friande bevis.

I studien som beskrivs i kapitel 4 undersöktes effekterna av den ordning i vilken bevismaterialet presenterades. Förväntningen var att det först presenterade bevismaterialet skulle påverka efterföljande beaktande av bevismaterial. I motsats verkade uppfattningen om den åklagades sannolika skuld påverkas mest av det bevismaterial som presenterades sist för deltagarna. Emellertid sträckte sig inte effekten på skulduppfattningen till att omfatta andelen fällande domar. Sammantaget bedömde deltagarna, som var juridikstudenter, att bevismaterial som tenderade att fria den åtalade var signifikant viktigare än det fällande bevismaterialet, vilket var motsatsen till vad som förväntades.

I studien som redovisas i kapitel 5 testades effekten av olika instruktioner om hur ett beslut rörande straffrättslig skyldighet skulle motiveras. För att göra detta skapades fyra uppsättningar med instruktioner. Dessa var baserade på den nederländska straffprocessordningen, den tyska straffprocessordningen och falsifieringsprincipen. Slutligen inkluderades en kontrollgrupp som gavs minsta möjliga instruktion. Deltagarna gavs en fallskiss som tydde på den tilltalades skuld. På grund av betoningen på att ta alternativa förklaringar i beaktande i den tyska straffprocessordningen och i falsifieringsprincipen, förväntades

dessa instruktioner leda till en lägre andel fällande utslag jämfört med de andra. De olika grupperna uppvisade emellertid inte någon skillnad i andel fällande utslag. Icke desto mindre inkluderade deltagare i den grupp som instruerades av den tyska straffprocessordningen och falsifieringsprinciperna signifikant större mängd friande bevismaterial i sitt skriftliga avgörande än deltagarna i kontrollgruppen gjorde. Enligt dessa resultat förefaller inverkan av olika instruktioner kring domarens ansvarsskyldighet påverka vilket bevismaterial som inkluderas i det skriftliga avgörandet, men inte det slutliga friande eller fällande beslutet. Detta skulle kunna betyda att det finns en diskrepans mellan det bevismaterial som deltagarna tog i beaktande och det bevismaterial som deltagarna inkluderade i det skriftliga avgörandet.

I den sista experimentella studien (kapitel 6) utforskades huruvida Analysis of Competing Hypotheses (ACH) – en teknik för strukturerat övervägande av scenarier – kan leda till att friande bevismaterial och alternativa scenarier tas i beaktande i större utsträckning. Deltagarna, som återigen var juridikstudenter, fick en fallskiss som beskrev en tydlig huvudmisstänkt men också innehöll en antydan om ett alternativt scenario. I motsats till förväntningarna gav deltagare i både kontroll- och ACH-gruppen företräde åt det friande bevismaterialet framför information som tydde på den huvudmisstänktes skuld. Det förefaller vidare som om många av deltagarna inte använde ACH korrekt, vilket stöder resultat från andra forskare som har antytt att ACH kan vara alltför vagt beskrivet för att kunna användas korrekt.

Resultaten från alla studierna sammanvägs och diskuteras i kapitel 7. Den huvudsakliga slutsatsen av resultaten är att domare är medvetna om risken för konfirmeringsbias och vikten av falsifiering och beaktande av alternativa scenarier. De tycks emellertid ha svårighet att använda sin kunskap i praktiken. En därmed förknippad slutsats är att det tycks finnas stora variationer i hur domare uppfattar sin aktiva roll, vilket skulle kunna vara ett område där användning av falsifiering kunde uppmuntras och förbättras. Slutligen tyder juridikstudenters observerade preferens för friande bevismaterial på en

diskrepans mellan juridikstudenters beslutsfattande i experimentsituationer och domares praktiska erfarenhet, även om användandet av juridikstudenter kan ha begränsat värdet av resultaten på två av de experimentella studierna. En annan slutsats kan därför vara att domare begränsas i sin användning av falsifiering av systemiska hinder som ligger utanför deras kontroll. Denna slutsats stämmer också överens med observationen att domare verkar förstå teorierna bakom falsifiering och alternativa scenarier, men har svårighet att använda dessa teorier korrekt i praktiken.

Samenvatting (Dutch Summary)

Hoewel er verschillende juridische regels zijn over de toelaatbaarheid van bewijs en het minimum bewijs dat nodig is, is er weinig bekend over hoe bewijs wordt gewogen en wordt geïntegreerd om tot een beslissing te komen over of de verdachte het feit heeft gepleegd waarvan hij wordt beschuldigd. De tenlastelegging kan gezien worden als de theorie dat de verdachte schuldig is. Een belangrijke stap in het uitzoeken of een theorie juist is, is op zoek gaan naar bewijs dat de theorie tegenspreekt. Die stap wordt ook falsificatie genoemd. Falsificatie is ook gerelateerd aan de scenariomethode, waarbij alternatieve scenario's worden vergeleken om het scenario te vinden dat het bewijs het best kan verklaren. De overweging van alternatieve scenario's kan worden belemmerd als de rechter een voorkeur heeft voor bewijs dat de theorie van de tenlastelegging ondersteunt en minder aandacht besteedt aan bewijs dat die theorie tegenspreekt. Verschillende onderzoekers hebben steun gevonden voor die neiging, bekend als confirmation bias, en de mogelijke invloed daarvan op het gedrag en beslissingen in verschillende werkvelden, inclusief politieonderzoeken en rechterlijke beslissingen.

In het huidige proefschrift is de rol van falsificatie in rechterlijke beslissingen onderzocht. Nadat een overzicht is gegeven van de bestaande theorieën over rechterlijke beslissingen in Hoofdstuk 1, worden er twee studies gerapporteerd, waarin het gebruik van falsificatie door rechters in de praktijk werd onderzocht (Hoofdstukken 2 en 3). Verder zijn er drie experimentele studies uitgevoerd om manieren te vinden waarop falsificatie zou kunnen worden gefaciliteerd (Hoofdstukken 4-6).

De studie die is beschreven in Hoofdstuk 2 gaat over rechterlijke beslissingen in een moordzaak. De veroordeling werd geanalyseerd op pogingen tot bevestiging en falsificatie op basis van het bewijs in het dossier. Gebaseerd op die analyse, lijkt het zo dat aan bewijs dat de schuld van de verdachte ondersteunt onevenredig meer waarde werd toegekend dan aan bewijs dat de schuld van een alternatieve verdachte ondersteunt, of aan bewijs dat de schuld van de verdachte

ontkracht. In het vonnis stonden ook verschillende alternatieve scenario's die de rechtbank had overwogen. Het lijkt er daarom op dat de rechtbank heeft gepoogd om alternatieve scenario's te overwegen, maar moeite had om dat effectief en in lijn met de theoretische onderbouwing van de scenariomethode uit te voeren.

In Hoofdstuk 3 werden de uitkomsten van een vragenlijst die is afgenomen bij strafrechters beschreven. Met de vragenlijst werd inzicht verkregen in het begrip van rechters over falsificatie en in hoeverre zij het nodig vinden om falsificatie te gebruiken. Daarnaast beschreven de rechters ook hoe zij falsificatie toepassen in de praktijk. Vanwege een laag aantal respondenten zijn de antwoorden op de vragenlijst aangevuld met informatie uit interviews met strafrechters. De bevindingen van dit onderzoek waren vergelijkbaar met de bevindingen van de studie uit Hoofdstuk 2. Respondenten hadden een goede theoretische kennis van falsificatie, maar vonden het lastig om die kennis toe te passen in de praktijk. Het werd ook duidelijk dat er veel variatie is in hoe rechters hun eigen rol zien en de mate waarin zij het nodig vinden om zelf alternatieve scenario's en ontlastend bewijs te overwegen.

In het onderzoek dat wordt beschreven in Hoofdstuk 4, werd de invloed onderzocht van de volgorde waarin het bewijs wordt gepresenteerd. De verwachting was dat het eerste bewijsstuk de overweging van daaropvolgend bewijs zou beïnvloeden. In tegenstelling tot die verwachting, werd de beoordeling van de waarschijnlijkheid dat de verdachte schuldig was vooral beïnvloed door het laatste bewijsstuk dat de proefpersonen te zien kregen. Dat effect was niet te zien in de beslissing om te veroordelen. De proefpersonen, die rechtenstudenten waren, beoordeelden ontlastend bewijs als significant belangrijker dan het belastende bewijs. Dat was tegen de verwachting in.

Hoofdstuk 5 handelt over een studie waarin het effect van verschillende instructies voor het uitleggen van een beslissing is onderzocht. Er waren vier soorten instructies gecreëerd. Drie daarvan waren gebaseerd op het Nederlands wetboek van Strafprocesrecht, het Duitse wetboek van Strafrecht, en het concept

van falsificatie. In de vierde conditie, die werd gebruikt als controleconditie, werd enkel gevraagd om de beslissing uit te leggen, zonder daarvoor verdere instructies te geven. Proefpersonen kregen een korte beschrijving van een misdrijf (vignette) te lezen dat overwegend belastend was voor de verdachte. Vanwege de nadruk op het overwegen van verschillende verklaringen in het Duitse wetboek van Strafrecht en bij falsificeren, werd verwacht dat het aantal veroordelingen lager zou zijn in die condities dan in de andere condities. Er werd geen verschil gevonden tussen de condities voor de veroordelingen. Toch gebruikten de proefpersonen in de Duitse conditie en de falsificatieconditie significant meer ontlastend bewijs in hun motivering dan de proefpersonen in de controleconditie. Gebaseerd op deze bevindingen lijken verschillende instructies om een beslissing te verantwoorden een invloed te hebben op het bewijs dat wordt gebruikt in de motivering, maar niet op de beslissing over schuld. Dat zou kunnen wijzen op een discrepantie tussen het bewijs dat wordt overwogen en het bewijs dat wordt gebruikt in de motivering.

In de laatste experimentele studie (Hoofdstuk 6) is onderzocht of de Analysis of Competing Hypotheses (ACH) – een techniek voor de gestructureerde overweging van scenario's – ervoor kan zorgen dat ontlastend bewijs, of alternatieve scenario's, meer worden overwogen. Proefpersonen waren opnieuw rechtenstudenten. Zij kregen een vignette te lezen waarin een duidelijk hoofdverdachte werd beschreven in een moordzaak, maar waarin ook een alternatief scenario werd gesuggereerd. In tegenstelling tot de verwachtingen hadden proefpersonen in zowel de controle- als de ACH-conditie een voorkeur voor bewijs dat ontlastend was voor de hoofdverdachte. Daarnaast bleek een groot deel van de proefpersonen ACH niet correct te hebben toegepast, wat overeenkomt met de bevinding van eerder onderzoek naar ACH, waarin uitgelegd werd dat ACH mogelijk te vaag is om naar behoren te worden toegepast.

De resultaten van alle onderzoeken worden samengevoegd en besproken in Hoofdstuk 7. De voornaamste conclusie op basis van de bevindingen uit de

beschreven onderzoeken is dat rechters zich bewust zijn van het gevaar van confirmation bias, en van het belang van falsificatie en alternatieve scenario's. Toch lijken ze het moeilijk te hebben om die kennis toe te passen in de praktijk. Een gerelateerde implicatie is dat er veel variatie is in hoe rechters hun actieve rol zien, wat de ruimte kan bieden om de toepassing van falsificatie aan te moedigen en te verbeteren. Tot slot suggereert de voorkeur voor ontlastend bewijs door rechtenstudenten een verschil tussen het beslissingsproces van rechtenstudenten onder experimentele omstandigheden, en de ervaring van rechters in de praktijk. Het is mogelijk dat rechters beperkt worden in hun gebruik van falsificatie door factoren in het rechtssysteem waar zij geen controle over hebben. Dat zou ook in overeenstemming zijn met de bevinding dat rechters theorieën over falsificatie en alternatieve scenario's begrijpen maar moeilijkheden hebben met ze naar behoren toe te passen in de praktijk.

Acknowledgements

Let me start by saying that I am bound to forget people in this chapter, or to leave out memorable experiences that I will forever remember, but would require a book of their own to be explained properly.

Starting with my supervisory team: thank you! Karl, Peter, and Robert, all of you have made major contributions to this thesis and the process leading up to it. Karl, you are the best supervisor anyone could ask for. Although you are very critical in your comments and feedback, you have perfected the art of sugar-coating to soften the blow. Besides the constructive feedback, it has been great to work with you on designing and executing studies. In a way, I wish my time in Gothenburg had been longer, because you were a pivotal factor in making the last three years a learning experience. I hope to have the honour to work with you again in the future.

Although working with Peter is a privilege, it was also the most daunting experience. Your feedback never failed to strike a chord. Over the years I noticed your feedback became increasingly positive, although I still have to read every positive email twice to believe it. I'm going to choose to interpret that as a sign of the progress I have made as a researcher over these past years, thanks to your input and help. I believe you have taught me some of the most important lessons that I will carry forward in whatever I end up doing in the future.

Robert, without you I would not have stayed in Maastricht, and I would never have applied for the House. You have been a key part of this journey and have fulfilled so many roles of support, from dealing with difficult master students to dealing with difficult promotors. Without the countless rants you listened to over over-priced coffee or wine, this thesis would probably never have been written. Besides working out research ideas with you from start to finish, our walks also served as history lessons for this town. Thank you for everything, including your continued support as I try to figure out what's next!

From supervisors onto the people I have shared my Maastricht supervisors with. Of course, Nikky and Ale, who have been there throughout the process.

Over several bottles of wine, many frustrations have been shared, too often without a solution. But somehow, we all made it to the end and we have beautiful memories to look back on. Besides conferences, there were our trips to the Efteling, Kruller Muller, and Budapest. I have also learned from people who previously had my supervisors. Tanja and Ricardo, you both shared your wisdom and experience, which has helped me more than I think you know. Both of you also helped me gain the confidence to pursue this path in the first place. Thank you for giving me perspective!

I have also been fortunate enough to be part of a program, meaning I was never alone in my struggles or successes. Of course, the Maastricht people were closest to me. Kathy, thank you for being supportive, welcoming, and kind, in every way that I can think of. Thank you for getting me travelling to see some beautiful parts of Europe that I might otherwise never have visited. I hope our paths will cross again soon! Irena, you are one of the kindest people I know, although that is often clouded by your honesty. Nevertheless, a refreshing experience I am grateful for. Bri, I have never quite understood how you could make me laugh about the silliest things, but I am grateful for the many random giggles with you, often much needed at stressful times. Nkansah, Renan, Nina, Sergeii, and Carey - thank you for being friends rather than colleagues! Of course, I am also grateful to the House of Legal Psychology for the funding that enabled me to take on this project. Thank you to all the people who made that possible, and to Chantal, who keeps it all running smoothly. The research I was able to conduct after the last few years also required some help from eager and capable students, without whom my life would have been a lot more difficult. Thank you Paulien, Imke, Carina, and Femke!

Although my interaction with most of the department members at the Law Faculty was unfortunately often limited to the annual outings, I am nevertheless grateful for the community that you all create, and the support that you have given me in my research and other endeavours. For that, I would like to specifically thank Miet, who made me feel welcome in the beginning and

continues to be a pleasure to work with, both due to your academic support as well as your sincere personality. Violet, as a former colleague but continued friend, there have been many conversations about shared interests, that I am sure we can continue in the future. To everyone I met in Portsmouth, to all the members of HAC lab, thank you for getting me through my first mobility. Especially Char, who I had expected to be my drinking buddy, but ended up being my gym buddy instead. To everyone in CLIP, thank you for being so open and welcoming to those just passing through. Meghana, Bruna, and Jenn – safe to say my time in Gothenburg would not have been the same without you, and I never expected to become as close as we did during that time. Special thanks to Jenn for being my running buddy and keeping me going!

At times when I started losing motivation about my research, I could always turn to Steef to get me riled up about all the problems with criminal law procedures, or anything else really. The outrage and passion with which we discussed such important matters, only to fall back in line the next day, were some of the main driving forces throughout this process. Living together has definitely created some new frustrations for us to be passionate about, but I have faith that we will never lose sight of what we mean to each other. Thank you for your help with the materials, the writing, the frustrations, and for being there on the day that it all comes down to.

A special thanks also goes to the friends who were further removed from work. Ele, although we have now been living in different countries for longer than we ever lived in the same country, coming to see you still feels like coming home. I have missed you so many times, and our little escapes from reality have been key in keeping me sane. Thank you for being a safe haven and for always being there! The same also goes for Laura. At the end of our time at RHUL, I referred to you as my voice of reason. That is still true. We have both come a long way in the past 6 years, and I am so grateful that we have continued to be there to witness the progress we have made. Your visits to Maastricht and Gothenburg, as well as opening up your house to me during my time in England, meant a lot

to me. Although you didn't always agree, you were always there. And thank you for designing the cover that was the finishing touch to this journey!

The last category of people I have to thank is family. To my parents and their partners, thank you for always being there and welcoming me whenever I needed to get away. Your unwavering support throughout life has gotten me to this point, and I am grateful for all the chances you have provided me with. To my grandparents, each of you have contributed in your own way to making me who I am today, for which I am grateful. Life lessons learned are carried forward in memories that live on. To my siblings: Gwijde, I think we both know that without our (un)healthy dose of sibling rivalry, I may not have ended up here. Although you may be my big brother, you were right beside me during this entire process and in a way, I feel we have become closer. The fact that we could go to each other with questions about writing or stats has been meaningful, and has perhaps bridged the English Channel that has separated us. As we both start new chapters, I wish you all the best in whatever path you choose. I hope that in the future, our paths will not only run parallel, but will cross more often than they have up to this point. To Töpi, thank you for being the energetic light point in my life. Your view on the world is still refreshing and always makes for a great distraction. I hope that hasn't changed by the time you can read this, but I somehow think your strong character will persevere through any obstacles you may face. I wish you the same support, love, and kindness from the people around you that I have had the honour of experiencing.

Research Dissemination

Books

Maegherman, E.F.L., De Boer, M.M., Horselenberg, R., & Van Koppen, P.J. (2020). *De onzichtbare steekpartij.* Boom Juridisch.

Published or Accepted Journal Articles

Maegherman, E., Horselenberg, R., & Van Koppen, P.J. (2020). Testing the Analysis of Competing Hypotheses in legal decision-making, *Applied Cognitive Psychology.*

Maegherman, E., Horselenberg, R., Ask, K., & Van Koppen, P.J. (2019). From pervasive beliefs to wrongful convictions, *The inquisitive mind*, 1.

Maegherman, E., Ask, K., Horselenberg, R., & Van Koppen, P.J. (2020). Law and order effects: On cognitive dissonance, belief perseverance, and confirmation bias. *Psychiatry, Psychology, and Law*

Maegherman, E., Van Veldhuizen, T., & Horselenberg, R. (2018). Dropping the anchor: the use of plausibility in credibility assessment. The Oxford Monitor of Forced Migration, 7(2), 37-55.

Sauerland, M., Otgaar, H., **Maegherman, E.,** & Sagana, A. (2020). Allegiance bias in statement reliability is not eliminated by falsification instructions. Zeitschrift für Psychologie, 228 (3), 210-215.

Van Veldhuizen, T.S., **Maegherman, E.F.L**., & Horselenberg, R. (2019). Het verhaal gewogen: Kennis van strafrechters over het geheugen en politieverhoor, Strafblad, 4, 58-63.

Horselenberg, R., & **Maegherman, E.** (2017). Valkuilen van verhoren bij en door de rechter. Strafblad, 2, 183-188.

Other Publications

Maegherman, E., & Vanderhallen, M. (2018). Effective Police Station Legal Advice - Country Report 1: Belgium. *Project Report.* University of Nottingham, Nottingham.

Maegherman, E., & Vanderhallen, M. (2018). Effective Police Station Legal Advice - Country Report 4: The Netherlands. *Project Report*. University of Nottingham, Nottingham.

Submitted Manuscripts

Maegherman, E., Ask K., Horselenberg, R., & Van Koppen, P.J. (2020). *Accountability and falsification in legal decision-making*. Manuscript submitted for publication.

Conference Presentations

Maegherman, E., Horselenberg, R., Ask, K., & Van Koppen, P.J. (July 2019). Popper in practice: Do judges use falsification? Oral presentation at the European Association of Psychology and Law conference, Santiago de Compostela, Spain.

Maegherman, E., Horselenberg, R., Ask, K., & Van Koppen, P.J. (June 2018). On the relation between cognitive dissonance, belief perseverance, and order effects. Poster presentation at the European Association of Psychology and Law conference,Turku, Finland.

Maegherman, E., De Boer, M.M. Horselenberg, R., & Van Koppen, P.J. (May 2017). Murder during carnival: Drunk eyewitnesses. Oral presentation at the European Association of Psychology and Law conference, Mechelen, Belgium.

Maegherman, E., Nieuwkamp, R., & Horselenberg, R. (May 2017) Knowledge evidence in alibis. Oral presentation at the European Association of Psychology and Law conference, Mechelen,Belgium.

Maegherman, E. & Vanderhallen, M. (May 2017) Effective police station legal advice in the Netherlands and Belgium. Oral presentation at the European Association of Psychology and Law conference, Mechelen, Belgium.

Curriculum Vitae

Enide Maegherman was born on July 29th, 1993, in Belgium. After moving to Budapest in 2011, she completed the International Baccalaureate at the British International School of Budapest. She then studied psychology at Royal Holloway, University of London, achieving a First Class Honours Bachelor degree of Science. Besides her interest in Psychology, she also had a keen interest in law, and subsequently moved to Maastricht to complete the Masters in Psychology and Law at the faculty of Psychology and Neuroscience. After obtaining that MSc. Degree (Cum Laude), she started the LL.M. in Forensics, Criminology, and Law at the Law Faculty of Maastricht University. She graduated Cum Laude in August 2016, and started working at the Law Faculty as a research and teaching assistant in September 2016. After working on several research projects and courses, she started her PhD with the House of Legal Psychology in September 2017. Her research project was conducted at Maastricht University under the supervision of professor dr. Peter van Koppen and dr. Robert Horselenberg, and at Gothenburg University under the supervision of prof. dr. Karl Ask. She spent six months in Portsmouth and six months in Gothenburg for her mobility periods. In May 2020, she submitted her thesis to Gothenburg Univerisity, followed by Maastricht University. Following her defence in January 2021, she hopes to conduct research in an applied setting, combined with the application of research to practice through training formats or by applying relevant knowledge to cases.

Appendix A.

Material for Chapter 3 (Translated from Dutch)

Part 1

For each of the statements below, please indicate whether you agree or disagree with the statement.

1) If there is enough convincing evidence for the guilt of the suspect, the police has conducted enough investigation. Excluding other scenarios is not necessary.
   - o Agree
   - o Disagree

2) Evidence for the suspect's innocence does not have to be explained as long as there is sufficient convincing evidence for the guilt of the suspect.
   - o Agree
   - o Disagree

3) The evidence, that supports the findings of the court that the charges have been proven and is mentioned in the reasoned decision, is the only evidence that was considered.
   - o Agree
   - o Disagree

4) Miscarriages of justice are caused by the (individual) evidence (e.g. wrongful identification, false confession).
   - o Agree
   - o Disagree

5) When reading the case file, a considered analysis has to be made of the evidence that contradicts the charges.
   - o Agree
   - o Disagree

6) If there is enough evidence that supports the guilt of the suspect, I do not doubt my belief in the suspect's guilt, despite the presence of evidence that supports the suspect's innocence.
   - o Agree
   - o Disagree

7) The aim of the police investigation is to collect information so that a correct legal decision concerning the evidence can be made.
   - o Agree
   - o Disagree

8) It is important that the police has investigated alternative scenarios. That should be visible from the case file.
   - o Agree
   - o Disagree

9) The consideration of alternative scenarios (which are apparent from the case file or as put forward by the defence) is up to the judge.
   - o Agree
   - o Disagree

10) It would be useful to, in addition to the case file, also study the explicit arguments made by the defence lawyer prior to trial.
   - o Agree
   - o Disagree

11) If a suspect confessed, the rest of the evidence should still be looked at critically (despite article 359 SV).
   - o Agree
   - o Disagree

12) There would be added value in also showing consideration of evidence that contradicts the charges in the written motivation for the judge's decision.
    - o Agree
    - o Disagree

If you have any questions or comments about the statements above, please write them down here. If not, you can simply continue the survey without filling anything in.

Part 2

The second part of the survey consists of 3 small tasks. Please provide what you think is the correct answer.

1) You want to know whether the following statement is true:
   "If a plant has round leaves, then it is in a blue pot"

   In order to check whether this statement is true, there are four plants you can examine.
   - A plant with round leaves
   - A plant with pointy leaves
   - A plant in a blue pot
   - A plant in a red pot

   Which of these plants should you examine to determine whether or not the statement is true. Examine as few plants as possible.
   - o The plant with the round leaves
   - o The plant with the pointy leaves
   - o The plant in the blue pot
   - o The plant in the red pot

Can you please explain your choice?

2) You are visiting a friend. Your friend's son says he is going to take a bath. A little while later, you notice there is a leak in the kitchen. You suspect that the son has spilled water during the bath. There are several questions you could ask.
   a. Is the bathroom above the kitchen?
   b. Is it raining?
   c. Was it already leaking before the boy went upstairs?

What question do you want to have answered in order to test your hypothesis? Can you please explain your choice?

3) On the 15th of November 2015, around 23.30, there was a fire in a shop on the ground floor. Above the shop, there were several floors with apartments.

   At the time of the fire, no one was present in the shop. However, in floors above the shop, people were present. As a consequence of the fire, four people had to be taken to the hospital with respiratory problems.
   The police started an investigation. It became clear that the fire was caused by arson. A suspect was arrested by the police. The suspect had worked at the shop for a short period of time. He had an earlier conviction for arson.

   Witness Johnson states he saw who started the fire. He said he saw the perpetrator from a short distance. The perpetrator was stood underneath a street light, so witness Johnson had a good view of the perpetrator. Witness Parker was stood next to witness Johnson. She also had a good view of the perpetrator. Two other witnesses, Adams and Baker, saw the perpetrator run away from the crime scene. They saw him come out of the shop and were walking right behind the perpetrator, which means they could see him well. The four witnesses were presented with a multiple photo line-up,

which included the photo of the suspect. All line-ups met the requirements of a properly conducted line-up.

Withness Johnson said he was sure that none of the people in the photos were the perpetrator. Witness Parker identified the suspect as the perpetrator. Witness Adams said none of the people in the line-up were the perpetrator. Witness Baker said he could not see the perpetrator in the line-up.

Based on the evidence above, would you convict or acquit the suspect?

- o Convict
- o Acquit

What is your verdict based on?

Part 3

You have now reached the final section of the survey. In this section, you will be asked a number of open questions. Please try to be as concrete as possible in your answers.

According to you, what is falsification?

How do you attempt to falsify the charges?

Can you give an example from practice?

How do you try to verify the charges?

Can you give an example from practice?

Do you think it is important to attempt falsification?

- o Yes
- o No

Can you please provide arguments for your choice?

Do you feel able to falsify in every case file?

- o Yes
- o No

If so, what do you need in order to apply falsification?

If not, what do you feel is missing in order to apply falsification?

To what extent do you try to find things out yourself (think for example of investigating alternative scenarios, testing findings, or collecting evidence outside of or during the trial)

What is the most common reason for you to want to do so?

Appendix B.
Materials for Chapter 4

Case Vignette

The Event

At 02:21 on a Sunday morning, the emergency central receives a call from a woman (Witness A) who reports having seen a taxi driver probably being shot outside her apartment. A police patrol and an ambulance are sent to the given address. In the taxi, which is parked in a lit up turning space, the police find a dead man in the driver's seat, with a shot wound in the head. Reinforcements are requested and the scene is cordoned off.

Witness A, who called the emergency central, lives on the first floor with her bedroom in the end of the house, facing the turning space. At 02:15 she was woken up by an aggressive male voice coming from the turning space, but she could not hear what was being said. After a minute or so she heard a bang, after which she walked to the window to see what had happened. At the turning space she saw the taxi, and since the interior lighting of the taxi was on she could see the driver hanging over the wheel. Then she called 112. In the outskirts of the residential area, a police officer on the way to the scene noticed a man running. The man was arrested and brought in for questioning.

Investigation Results

According to the taxi company, a customer called in and ordered a taxi to the turning space at the specified time. The customer, however, has a reliable alibi and has no involvement in the crime.

At the crime scene, the police found shoeprints around the car.

The post-mortem examination showed that the victim had died as a result of a shot to the head. The bullet had entered slightly above the victim's right eye and had been fired from a very short distance (< 1 meter). Both the fired bullet and a 9 mm cartridge were found inside the taxi.

The police's canine squad also found a pair of leather gloves and a 9 mm pistol of the make Zastava in a grove near the place where the suspect was arrested.

The results from the National Laboratory of Forensic Science showed that the bullet and the cartridge had been fired from the discovered gun, and that the leather gloves showed traces of powder stain.

A check of the taxi's daily receipts showed that €110 in cash was missing from the car. It also appeared as if a GPS and a mobile phone were missing.

The Suspect

The suspect, who had been seen running in the outskirts of the residential area, was identified using his driver's license.

Records showed that he had been previously convicted of aggravated assault and illegal carrying of a knife. In the first police interview, the suspect claimed that he was running to get away from three men with whom he had been in a fight earlier that night.

When asked to explain which way he had been running, the man answered that he had passed the turning space and that he recalled a taxi being parked there. After that, the suspect refused to answer any further questions.

The suspect lives in the vicinity of the crime scene.

The man is 27 years old, 187 cm tall, and weighs 79 kg. At the time of the arrest he was dressed in dark-green, baggy trousers, a black leather jacket, and a light-gray hooded sweater.

Additional Evidence

Type of Evidence	INCRIMINATING	EXONERATING
Eyewitness	The police have found an eyewitness who saw someone leaving the taxi area around the time of the crime. The witness was presented with a line-up at the police station the following day and identified the suspect as the perpetrator. The witness was confident in his decision.	The police have found an eyewitness who saw someone leaving the taxi area around the time of the crime. The witness was presented with a line-up containing the suspect at the police station the following day. The witness was confident that the perpetrator was not present in the line-up.
Hair	The hair found in the glove near the crime scene was sent to the National Laboratory of Forensic Science for close comparison to the hair of the suspect. The lead scientist on the case has declared that the hair from the crime scene very probably originated from the suspect (85% certainty).	The hair found in the glove near the crime scene was sent to the National Laboratory of Forensic Science for close comparison to the hair of the suspect. The lead scientist on the case has declared that the hair from the crime scene very probably did not originate from the suspect (85% certainty).
CCTV	After obtaining a warrant the police are allowed to inspect the CCTV footage from the turning space. Close observation of the footage shows that someone who matches the physical appearance and clothing of the suspect was at the turning space around the time of the crime.	After obtaining a warrant the police are allowed to inspect the CCTV footage from a bar a few blocks away from the crime scene. Close observation of the footage shows that someone who matches the physical appearance and clothing of the suspect was at the bar at the time of the crime.
Shoeprints	The suspect's shoes were investigated to see whether they matched the shoeprints found around the taxi. The shoeprints matched the shoes the suspect was wearing when he was arrested.	The suspect's shoes were investigated to see whether they matched the shoeprints found around the taxi. The shoeprints did not match the shoes the suspect was wearing when he was arrested.

Appendix C.
Supplemental Materials to Chapter 4

Material

As there were eight pieces of evidence in total, and each condition used only four pieces, there were many possible combinations of evidence which could have come about if participants were randomly assigned pieces of evidence during the survey. It was therefore decided to create four sub-conditions within each condition. Each of these sub-conditions presented a different combination of evidence types in the order dictated by the condition. Each piece of evidence occurred equally often in the sub-conditions.

The evidence that was used in the different conditions was pretested in nine pretests to find pieces of evidence that were approximately equal in strength for their incriminating and exonerating counterparts. In the final prestest (N = 78; with the exception of shoe prints which was added later, N = 35), the average strength ratings for the different pieces of evidence were all within 10 points from their counterparts on a 100 point scale (see Table S1).

Table S1

Average Strength Ratings (and Standard Deviations) for the Different Pieces of Evidence in the Final Pretest

Evidence	Exonerating	Incriminating
Witness	52.60 (5.47)	61.46 (4.26)
Hair	60.41 (4.97)	69.40 (8.56)
CCTV	59.27 (5.01)	51.28 (7.51)
Shoeprints	66.17 (11.22)	58.51 (9.62)

Language

To examine whether the survey language influenced the responses, independent *t*-tests were conducted on each of the response variables (Table S2).

Chi-square tests were conducted for the binary variables of whether participants would convict the suspect or not (Table S3). Alpha levels were adjusted to .003 after a Bonferroni correction was applied. None of the variables differed significantly between the Dutch and the English group. There was no interaction effect between language and condition for likelihood guilt between the different times, $F(5, 395) = 1.42$, $p = .215$, $\eta_p^2 = .018$, 90% CI [0.00, 0.03]. Language also did not interact with condition for cognitive dissonance between the different times, $F(5, 395) = 1.70$, $p = .135$, $\eta_p^2 = .021$, 90% CI [0.00, 0.04].

Table S2

Means (and Standard Deviations) on the Dependent Measures for Participants Taking the Study in Dutch and English

Dependent Measure	Dutch (n = 168)	English (n = 239)	t	p
Likelihood guilty (T1)	49.27 (19.01)	45.66 (21.22)	-1.50	.134
Confidence (T1)	59.88 (21.73)	64.15 (21.90)	2.47	.014
Cognitive Dissonance (T1)	2.07 (1.21)	2.29 (1.35)	1.72	.087
Likelihood guilty (T2)	48.82 (28.75)	49.01 (27.12)	0.21	.833
Confidence (T2)	66.65 (21.35)	66.09 (23.09)	1.20	.231
Cognitive Dissonance (T2)	2.00 (1.18)	2.33 (1.37)	2.09	.037
Likelihood guilty (T3)	47.60 (29.49)	49.42 (26.22)	0.51	.613
Confidence (T3)	69.14 (22.79)	63.89 (23.14)	0.03	.974
Cognitive Dissonance (T3)	2.13 (1.36)	2.59 (1.43)	2.18	.030
Importance Incriminating 1[a]	55.19 (25.39)	51.96 (28.28)	-1.06	.290
Importance incriminating 2[a]	55.76 (23.17)	52.12 (29.25)	-0.77	.444
Importance exonerating 1[b]	57.03 (26.05)	56.54 (28.01)	-0.47	.637
Importance exonerating 2[b]	61.81 (24.32)	59.72 (28.33)	-0.76	.451

Note. The Bonferroni-corrected significance level was set at .003. The importance ratings for the third and fourth pieces of exonerating and incriminating evidence, which were only included in the control conditions, could not be included as there were no valid pairs.

[a]Dutch (n = 139) and English (n = 204) as the incriminating evidence was not included in the exonerating control condition.

[b] Dutch (n = 137) and English (n = 206) as the exonerating evidence was not included in the incriminating control condition.

Table S3

Chi-square Tests for Differences Between Language Groups for the Binary Dependent Measures

Measure	χ^2	*p*
Convict (T1)	0.01	.915
Convict (T2)	1.22	.269
Convict (T3)	0.01	.946

Note. $N = 407$ and $df = 1$ for all analyses.

Table S4

Means (Standard Deviations) and Pearson Correlations for Change in Likelihood of Guilt (T2–T3) and Cognitive Dissonance (T3)

Condition	Change in likelihood of guilt (T2–T3)	Cognitive dissonance (T3)	*r*	*p*
Mixed Evidence				
Mixed 1	-0.29 (15.44)	2.65 (1.53)	-.127	.284
Mixed 2	-0.24 (14.81)	2.13 (1.23)	-.194	.116
Contradictory evidence				
Ex/Inc	25.24 (17.48)	2.79 (1.61)	.049	.687
Inc/Ex	-28.90 (21.79)	2.41 (1.40)	-.014	.909
Control conditions				
Ex/Ex	-11.05(16.34)	1.85 (1.20)	-.032	.799
Inc/Inc	10.83 (12.29)	1.91 (1.20)	.036	.776

Table S5

Mean Importance Ratings (and Standard Deviations) for the Different Pieces of Evidence

Condition	Evidence 1	Evidence 2	Evidence 3	Evidence 4
Mixed 1	44.73 (25.23)	59.14 (25.49)	52.63 (25.46)	63.59 (24.81)
Mixed 2	57.40 (25.64)	53.43 (24.34)	61.41 (26.60)	52.81 (27.93)
Ex/Inc	59.59 (27.48)	58.21 (27.15)	61.10 (25.65)	52.79 (25.62)
Inc/Ex	54.12 (30.37)	54.03 (28.23)	56.52 (27.12)	62.16 (24.42)

Note. The evidence number here refers to the order in which the evidence was received.

Appendix D.
Materials for Chapter 5

Instructions

Justification Condition

Please explain your decision on the guilt of the defendant.

- The decision should rest on the evidence that you mention in your verdict.
- Your verdict should include facts and circumstances that give reasons for your decision.
- If your decision differs from explicitly substantiated points raised by either the prosecution or defence**, give reasons for this.

** Points which the prosecution or defence provide evidence to support or prove the truth of

Explication Condition

Please explain your decision on the guilt of the defendant

- Your verdict should specify what relevant facts are deemed to be proven or not proven.
- Demonstrate that you considered and evaluated all relevant facts and circumstances both for and against your belief in judging the likelihood of the defendant's guilt.
- Explain any obvious alternative scenarios that are equally consistent with the facts as the scenario you decided on.
- Explain how you determined the weight of the individual pieces of evidence you considered.

Falsification Condition

Please explain your decision on the guilt of the defendant.

- Your verdict should describe the different possible versions of the events that you considered.
- Use the available evidence to explain how you excluded alternative scenarios.
- Explain how the evidence supports your decision to convict or acquit the defendant

Control Condition

Please explain your decision on the guilt of the defendant.

- Describe how you came to your decision.
- Your verdict should refer to the available evidence.

Case Vignette

On Monday the 23rd of January 2017, Emma Miller, James Miller's wife, found her deceased husband lying on his back on the bed in the bedroom of their suburban home. Upon her discovery, Emma called the emergency services and told the operator about what had just happened. After being informed by the operator, the police immediately rushed to the scene of the crime. When they arrived at the Miller home they found Emma covered in blood sitting next to her dead husband's body. It immediately became clear that James had multiple stab wounds in his chest.

Emma Miller was interviewed by the police. She claims to have left the house around 19:30 to visit her friend Catherine Hughes. Since James was visiting his parents and therefore not at home when she left, Emma claims she locked the front door to the house. Emma arrived at her friend's house around 20:30 but the police consider it suspicious that it took Emma an hour to get to her friend's house while this trip should normally only take her 30 minutes. According to Emma, she stopped by her office on the way, but this could not be confirmed. Emma claims to have left Catherine's house around 21:50 and arrived back home around 22:15. When she arrived home, she noticed the front door was unlocked.

When she called James' name but did not get a response, she decided to go look for him. This is when she found James dead on the bed they shared.

The police immediately start a large-scale investigation to clarify what happened to James Miller. Various pieces of forensic evidence were found during the investigation of the crime scene. The Technical Criminal Investigation Department found the victim lying in a pool of blood. Furthermore, they found bloody fingerprints on the edge of the bed that turned out to belong to Emma. They also found traces of blood on the wall behind the bed. On the pillows, they found both long brown and long blonde hairs. Emma's DNA was found at the crime scene and on James' body. DNA from an unidentified woman was found on the door handle of the bedroom door. In the bathroom sink, it was clear that Emma had washed her hands. The sink contained traces of James' blood and there was a bloody fingerprint on the tap. The fingerprint was Emma's. The police believe she was trying to wash away traces of evidence.

An autopsy of the victim's body confirmed that the stab wounds in the chest had been the cause of death. The stab wounds seemed to have been caused by a right-handed person, but the medical examiner was not certain about this. Time of death was between 19.30 and 20.30. It seems as if James had had sexual contact with a woman shortly before he died.

In order to find out who might have had a motive to kill James, the police start interviewing friends and family. Amongst the interviewees were two of James' friends: John Taylor and Paul Baker. John stated that James told him a few months ago that he was having an affair. Paul confirmed John's story and stated that James also told him about the affair, but about a week before James told John. Neither of the witnesses could confirm who the mysterious mistress is, but both testify that they had seen him talking to a brunette on Thursday January 19th in the bar where they always play darts. Judging from how James was communicating with her and gently touching her, they were under the impression that their friend and the unknown woman were intimate with each other. The only thing John and Paul can confirm is that the woman was not

Emma, as she has blonde hair, not brown. Eventually, police were unable to track James' presumed mistress down. According to John Taylor, James had been planning on ending the affair as soon as possible, because he could tell Emma was very suspicious. John thought James might have planned to meet his mistress on Monday night, as James had said he could not meet at the bar that night.

After finding out about James' mistress, the police now suspect Emma has killed her husband out of anger over the affair. Friends of Emma told the police that she had previously threatened to hurt James if he cheated on her. A few days ago, she had told one of her friends that she thought James was cheating on her and that she was looking for proof. Emma showed almost no emotion when talking to the police about James' death. According to the prenuptial agreement, Emma and James would divide their possessions equally in case of divorce, but if one of them died, the other would get everything. In order to get a clear timeline of the events that night, they decide to interview Emma's friend Catherine about the fatal night. Catherine confirms Emma's story completely and states that her friend was with her that night at the times indicated by Emma. However, through further investigation, the police find out that Catherine owes Emma a large sum of money and now believe that this is a valid reason for providing Emma with a false alibi for that night. Emma also received a parking ticket at 21:15 while being parked outside of her friend's house.

The police also interviewed the neighbours. One of them claimed to have heard screaming coming from the Millers' home somewhere between 15:00 and 17:00 on that specific Monday. According to this neighbour, it seemed as if a woman and a man had a fight, but she could not say if it were Emma and James she heard screaming. Emma was at home in the afternoon, which was confirmed by witnesses who saw her car in the driveway at 16.30. However, Emma denies fighting with James. Another neighbour heard someone arrive at the Millers' house that Monday evening between 19:45 and 20:00. He heard the front door close around that time but claims he did not hear the doorbell.

Another neighbour living a block further down, said that he saw a woman walking down the street that Monday evening. The woman was wearing a hood, so her face and hair were covered, but the neighbour claims the woman had Emma's posture. In his opinion, the woman seemed very nervous. This neighbour saw how this woman stopped briefly at one of the trash cans in the street and then disappeared from sight. Based on this story, the police searched the trash cans in the street. They found a carpet knife (as pictured below) which was covered in blood.

Based on DNA from the blood traces found on the knife, the Technical Criminal Investigation Department confirmed that this must be the knife used to kill James. Emma's fingerprints were found on the knife, and Emma is right-handed. There were also fingerprints which did not match James or Emma on the knife. It is not known who these prints belonged to.

In sum, the police believe Emma has killed her husband, based on the forensic evidence found at the scene, the fingerprints on the carpet knife, the witness testimonies from James' friends and the neighbours, her shaky alibi and the fact that she was found covered in blood at the crime scene. Emma is also the only person with a motive for killing James: the fact that he was having an affair. Emma's defence lawyer argues that the police should continue to try and find James' mistress.

Appendix E.
Supplemental Materials to Chapter 5

Coding

The evidence that participants used in their decision was coded according to 21 different categories. Based on the pre-test, five categories of evidence had been determined for the evidence in the case file: strongly incriminating (14), mildly incriminating (4), neutral (9), mildly exonerating (4), or strongly exonerating (3). The categories were determined based on the mean ratings of the evidence on a scale from 0 (*exonerating*) to 100 *(incriminating)* and their confidence intervals. The strongly exonerating category had a mean rating below 40, and the mildly exonerating category had a mean between 40 and 50 with an upper 95% CI bound no greater than 50. The neutral category had a mean between 40 and 60, with an upper 95% CI bound crossing 50. The mildly incriminating category had a mean between 50 and 60, and a lower 95% CI bound above 50. The strongly incriminating category had a mean rating above 60.

For each of the five categories determined by the pre-test, participants could use the evidence as incriminating, neutral, or exonerating, resulting in a total of 15 prespecified categories. Those 15 categories were then combined into the three categories that were used to test the hypotheses, namely incriminating, neutral, or exonerating evidence, as used by the participant

However, during the coding, it became clear that participants were sometimes not specific about which evidence they referred to. For instance, several participants mentioned that Emma had a 'motive'. As there were several pieces of evidence that related to a motive, such as the fact that Emma knew James was having an affair and that she would benefit financially more from his death than from a divorce, we decided to code the mention of such evidence as 'unspecified'. Some participants also misremembered information that was provided in the case. For instance, they mentioned that DNA was found on the murder weapon, whereas the case only specified that fingerprints were found on the murder weapon. Therefore, an additional category was created for "misremembered"

evidence. For both the unspecified and misremembered categories, a distinction was also made for whether the evidence was used as incriminating, neutral, or exonerating. Therefore, an additional six categories were created, resulting in a total of 21 categories to be coded. All the incriminating categories were combined into one incriminating category, which was used for the final analyses. The exonerating and neutral categories were also combined into one overall exonerating category, and one overall neutral category.

Valence ratings

Participants were asked to rate the valence of the evidence on a scale from 0 to 100, where 0 means completely exonerating and 100 means completely incriminating. The average of these ratings for each of the pieces of evidence is given in Table S1.

Table S1

Valence Ratings for the Evidence in the Case Vignette.

Evidence	*M*	*SD*	Category according to pre-test[a]
Emma was covered in blood when the police arrived at the crime scene	54.75	28.16	N
Emma's journey to Catherine's place took longer than usual	62.81	25.10	SI
Emma claimed to have stopped at work on her way to Catherine's place	46.79	25.90	N
Emma's bloody fingerprints were on the bed	49.80	29.07	N
Blonde hairs were found on the pillow	37.02	31.55	ME
Brown hairs were found on the pillow	39.14	31.48	ME
Emma's DNA was found at the crime scene	43.66	32.50	N
Emma's DNA was found on James' body	44.31	31.57	N
DNA of an unidentified woman was found on the door handle of the bedroom door	35.13	32.72	ME
The bathroom sink contained traces of James' blood	54.74	29.33	MI
Emma's bloody fingerprint was found on the bathroom tap	63.99	28.29	SI
The medical examiner thought the stab wounds had been caused by a right-handed person	53.47	26.49	MI
James died between 19.30 and 20.30	55.48	30.18	MI
According to the medical examiner, James had sexual contact with a woman shortly before he died	44.92	31.77	N
James had told his friends that he was having an affair	56.25	29.38	SI
James' friends thought his mistress was the brunette they had seen him with	43.73	30.70	N
According to John Taylor, James had been planning on ending the affair as soon as possible	39.28	29.10	ME

Table S1 (continued)

Evidence	*M*	*SD*	Category according to pre-test[a]
John Taylor thought James was meeting his mistress the night of the murder	42.40	31.46	N
According to her friends, Emma had threatened to hurt James if he cheated on her	71.82	23.98	SI
Emma had told her friend she suspected James was cheating on her	65.79	24.60	SI
Emma's friend claims Emma was looking for proof of an affair	64.96	25.46	SI
Emma showed almost no emotion when talking to the police about James' death	65.40	27.10	SI
Emma would benefit from James' death	71.81	23.92	SI
Catherine confirms Emma was with her at the times indicated by Emma	33.01	27.38	SE
Catherine owes Emma a large sum of money	63.87	24.33	SI
Emma received a parking ticket while parked outside Catherine's house	35.64	30.04	SE
A neighbour heard a fight in Emma and James' house earlier in the day	61.47	24.91	SI
Emma was home around the time of the fight	64.79	24.71	SI
Emma denies fighting with James that afternoon	56.85	26.67	MI
The neighbour heard someone arrive between 19.40 and 20.00, but didn't hear the doorbell ring	55.86	28.44	MI
A witness saw someone with Emma's posture stop briefly at the trash can where the knife was later found	66.61	23.75	SI
Emma's fingerprints were on the knife	69.57	28.18	SI
An unknown individual's fingerprints were found on the knife	29.89	31.09	SE
Emma is right-handed	60.07	27.25	SI

[a]*:SI = Strongly Incriminating, MI = Mildly Incriminating, N = Neutral, ME = Mildly Exonerating, SE = Strongly Exonerating.*

Appendix F.
Materials for Chapter 6

General Information about Biases Given to Both Groups

When reasoning, people can be influenced by cognitive biases. Cognitive biases are errors in judgment. They occur on a regular basis, and (almost) everyone shows them. These biases can have a negative effect on the way in which one tries to conduct an investigation or make a decision. One of the most common biases is the confirmation bias, which is a tendency to favour, and interpret, evidence in line with an existing belief. These biases can influence the interpretation and evaluation of evidence. For instance, one might pay less attention to evidence which is contrary to their initial belief.

ACH Information

When reasoning, people can be influenced by cognitive biases. Cognitive biases are errors in judgment. They occur on a regular basis, and (almost) everyone shows them. These biases can have a negative effect on the way in which one tries to conduct an investigation or make a decision. One of the most common biases is the confirmation bias, which is a tendency to favour, and interpret, evidence in line with an existing belief. These biases can influence the interpretation and evaluation of evidence. For instance, one might pay less attention to evidence which is contrary to their initial belief.

To counter this negative effect, several debiasing techniques have been invented over the years. One of these techniques is the Analysis of Competing Hypotheses. This method is used to evaluate and weigh alternative explanations and conclusions for a certain problem. It is a systematic way of making a decision based on different indicators for different hypotheses.

The ACH method has 3 characteristics. Firstly, at least two hypotheses should be formed. Secondly, ACH looks for evidence that is the most helpful in discriminating between the hypotheses, by either confirming or disconfirming

the hypotheses. The third characteristic is that ACH focuses on evidence that disconfirms a hypothesis, instead of evidence that supports it.

The ACH procedure consists of several steps. These steps are the core of this method. They are as follows:

1. Formulating multiple hypotheses: make a list of all the possible explanations for the problem you're facing.
2. Gathering evidence: make a list of all the evidence and arguments you have **for and against** your hypotheses.
3. Table construction: make a table with your hypotheses across the top (columns) and the evidence down the side (rows). Then evaluate the evidence in relation to the hypotheses in a row-by-row fashion. For each piece of evidence, determine whether it confirms or disconfirms the different hypotheses. Use signs (e.g., +, +/-,+/-- and -) to indicate if the evidence is confirming, neutral, or disconfirming with the hypothesis. Once the evidence has been evaluated for all hypotheses, you can move on to the next piece of evidence and repeat this step. Do this until you have analyzed all pieces of evidence.

Example:

	Hypothesis 1	Hypothesis 2	Hypothesis 3
Evidence 1	+	-	+
Evidence 2	+	+/-	-
Evidence 3	-	+	-
Evidence 4	+	+	-
Evidence 5	-	+	+

4. Draw conclusions: For each of the hypotheses, look at the amount of **disconfirming** evidence. The hypothesis with the least disconfirming

evidence is considered to be the most likely one. (In the example above, hypothesis 2 would be most likely)

Next, an example will be provided to illustrate the ACH method and to give you the opportunity to test your understanding of this method by constructing the table yourself. You will not be evaluated on your performance. The aim is simply to practice working with this method.

There has recently been a lot of burglaries in a suburban neighbourhood near Maastricht. The police do not have a lot of evidence, but there are three possible scenarios.

Let's assume we have the following hypotheses about the offender(s):

The burglar is a convicted burglar who has just been released from prison The burglar is the son of another burglar who was well-known to the police, but died last year. The burglaries were committed by a travelling group of foreign criminals

Now let's assume we have the following evidence:

The recent burglaries were committed using the same specific method previously used by the convicted burglar. The son of the well-known deceased burglar has just moved into this neighbourhood. Fingerprints found at the scene did not match the database. Some of the stolen objects were found at the pawn shop previously used by the well-known deceased burglar. The just released burglar has an alibi for 2 of the 7 burglaries. One witness overheard the burglar pick up the phone in regional Dutch dialect.

Now it's up to you to construct the table and to see which hypothesis is the most likely one.

Please try and use the ACH method during the following task.

Case

On Monday the 23rd of January 2017, Sabine Kostal, Jasper Kostal's wife, found her deceased husband lying on his back on the bed in the bedroom of their suburban home. Upon her discovery Sabine called 112 and told the operator about what had just happened. After being informed by the operator, the police immediately rushed to the scene of the crime. When they arrived at the Kostal home they found Sabine covered in blood sitting next to her dead husband's body. It immediately became clear that Jasper had multiple stab wounds to his chest.

Sabine Kostal was interviewed by the police. She claims to have left the house around 19:30 to visit her friend Catherine Huygens. Since Jasper was visiting his parents and therefore not at home when she left, Sabine claims she locked the front door to the house. Sabine arrived at her friend's house around 20:30 but the police consider it suspicious that it took Sabine an hour to get to her friend's house while this trip should normally only take her 30 minutes. Sabine claims to have left Catherine's house around 21:50 and arrived back home around 22:15. When she called his name, but did not get a response, she decided to go look for him. This is when she found Jasper dead on the bed they shared.

The police immediately start a large-scale investigation to clarify what happened to Jasper Kostal. Various pieces of forensic evidence were found during the investigation of the crime scene. The Technical Criminal Investigation Department found the victim lying in a pool of blood. Furthermore, they found bloody fingerprints on the edge of the bed that turned out to belong to Sabine. They also found traces of blood on the wall behind the bed and some hairs on the pillows. Sabine's DNA was found at the crime scene and on Jasper's body.

An autopsy of the victim's body confirmed that the stab wounds in the chest had been the cause of death. The stab wounds seemed to have been caused by a right-handed person, but the medical examiner was not certain about this. It seems as if Jasper had had sexual contact with a woman shortly before he died. Traces of lipstick were found on his body.

In order to find out who might have had a motive to kill Jasper, they start interviewing friends and family. Amongst the interviewees were two of Jasper's friends: Job Gerritsen and Paul Berings. Job stated that Jasper told him a few months ago that he was having an affair. Paul confirmed Job's story and states that Jasper also told him about the affair, but about a week before Jasper told Job. Both witnesses cannot confirm who the mysterious mistress is, but both testify that they had seen him talking to a brunette on Thursday January 19 in the café where they always play darts. Judging from how Jasper was communicating with her and gently touching her, they were under the impression that their friend and the unknown woman were intimate with each other. The only thing Job and Paul can confirm is that the woman was not Sabine, since the latter is a blonde, not a brunette. Eventually, police were unable to track Jasper's presumed mistress down.

After finding about Jasper's mistress, the police now suspect Sabine has killed her husband out of anger over the affair. In order to find out if Sabine knew about being side-tracked and to get a clear timeline of the events that night, they decide to interview Sabine's friend Catherine about the fatal night. Catherine confirms Sabine's story completely and states that her friend was with her that night at the times indicated by Sabine. However, through further investigation the police find out that Catherine owes Sabine a large sum of money and now believe that this is a valid reason for providing Sabine with a false alibi for that night.

The police also interviewed the neighbors. One of them claimed to have heard screaming coming from the Kostal home somewhere between 15:00 and 17:00 on that specific Monday. According to this neighbor, it seemed as if a woman and a man had a fight, but she could not say if it were Sabine and Jasper she heard screaming. Another neighbor heard someone arrive at the Kostal's house that Monday evening between 19:45 and 20:00. He heard the front door close around that time, but claims he did not hear the doorbell.

Another neighbor living a block further down, said that he saw a woman walking down the street that Monday evening. The woman was wearing a hood,

so her face and hair were covered, but the neighbor claims the woman had Sabine's posture. In his opinion, the woman seemed very nervous. This neighbor saw how this woman stopped briefly at one of the trash cans in the street and then disappeared from sight. Based on this story, the police searched the trash cans in the street. They found a carpet knife (as pictured below) which was covered in blood.

Based on DNA from the blood traces found on the knife, the Technical Criminal Investigation Department confirmed that this must be the knife used to kill Jasper. Next to that, Sabine's fingerprints were found on the knife.

In sum, based on the evidence collected, the police firmly believe Sabine has killed her husband. The forensic evidence found at the scene, the fingerprints on the carpet knife, the witness testimonies from Jasper's friends and the neighbors, her shaky alibi and the fact that she was found covered in blood at the crime scene, leave no doubt about her guilt. Sabine is also the only person with a motive for killing Jasper: the fact that he was having an affair.

Investigative Questions

- Is Sabine right-handed?
- Can anyone other than her friend confirm Sabine's alibi?
- Are there any indications that Sabine and Jasper recently had an argument?

- Was DNA from someone other than Sabine and Jasper found at the crime scene?
- Is Jasper's time of death between 19.30 and 20.30?
- Is there evidence that Sabine stopped somewhere on her way to her friend?
- Has Sabine ever expressed hostility towards Jasper?
- Was the front door locked when Sabine arrived back home?
- Has Sabine previously mentioned knowing about the affair to anyone?
- Did they find hairs on the pillow belonging to someone other than Sabine or Jasper?
- Did Sabine show any emotions in response to Jasper's death?
- Are there indications that Sabine was trying to save her marriage?
- Is there evidence that Sabine tried to get rid of forensic traces?
- Is there evidence that Jasper planned on meeting his lover that night?
- Are there indications that Sabine was at home in the afternoon the day of the murder?
- Is there reason to believe Sabine used the knife for something else?
- Does Sabine wear lipstick?
- Is there any evidence that Jasper was going to end the affair?
- Did Sabine admit to fighting with Jasper that afternoon?
- Did Jasper have contact, e.g. calls or texts, with his lover recently?
- According to the prenuptial agreement, would Sabine benefit from Jasper's death?
- Did they find fingerprints on the murder weapon belonging to anyone other than Jasper and Sabine?
- Has Jasper suggested to others that he was in love with his lover?
- Is there evidence that Sabine tried to do CPR on Jasper?

Appendix G.
Supplemental Materials to Chapter 6

Exclusions

Only two participants were excluded for having a background in something other than criminology or law. A further 82 participants were excluded for failing the attention checks. Finally, 86 participants were excluded for failing the control questions about the material. Those who conducted the study under the supervision of a researcher in the initial stages of data collection were not presented with the attention checks or control questions, and were therefore not subjected to the above exclusion criteria.

Language

Participants could choose to complete the survey in either English (n = 37) or Dutch (n = 154). Welch t-tests (with Bonferroni correction) were used to test for differences between the two language groups. No significant difference was found for any of the measures (Table S1).

Table S1

Welch t-Tests for Differences Between the Dutch and English Groups

Measure	*df*	*t*	*p*
Likelihood of guilt 1	47.5	0.666	.418
Likelihood of guilt 2	47.5	0.427	.516
Number of confirming questions	53.6	0.206	.652
Number of disconfirming questions	53.6	0.206	.652
Average importance rating of incriminating evidence	58.0	5.330	.025
Average importance rating of exonerating evidence	52.0	0.679	.414

Note. Bonferroni-corrected α = .008.

There was also no significant interaction effect between condition and language for any of the dependent variables, which was tested using 2 (condition) x 2 (language) ANOVAs (Table S2).

Table S2

Tests of the Condition × Language Interaction in Factorial ANOVAs for all Dependent Measures

Measure	*df*	*F*	*p*	η_p^2
Likelihood of guilt 1	1, 187	3.126	.079	.016
Likelihood of guilt 2	1, 187	0.825	.365	.004
Number of confirming questions	1, 187	0.415	.520	.002
Number of disconfirming questions	1, 187	0.415	.520	.002
Average importance rating of incriminating evidence	1, 168	0.397	.530	.002
Average importance rating of exonerating evidence	1, 183	0.292	.589	.002

Normality Tests

Based on Shapiro-Wilk tests, it was concluded that the data for the dependent variables were not normally distributed (Table S3). It was therefore decided to conduct non-parametric test for the main analyses.

Table S3

Normality Tests for the Dependent Measures

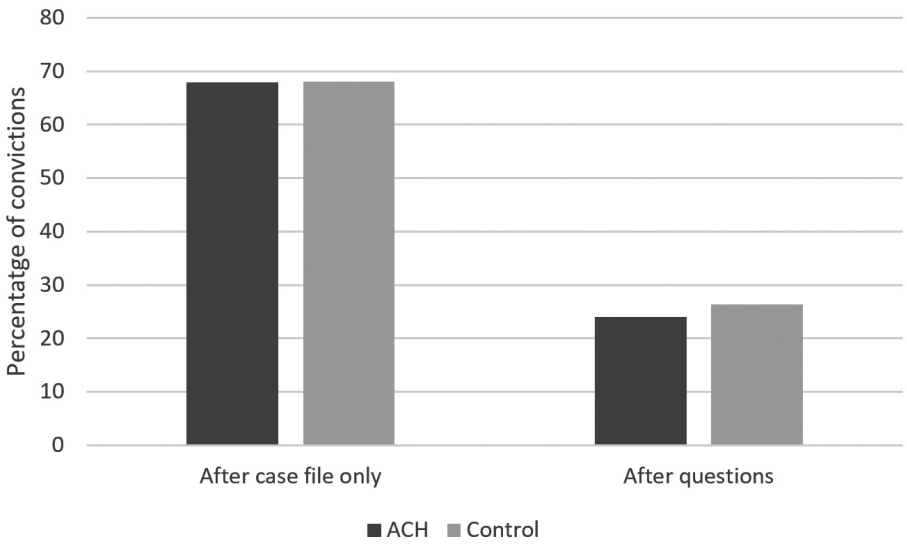
Measure	Shapiro-Wilk statistic	*df*	*p*
Likelihood of guilt 1	.972	172	.001
Likelihood of guilt 2	.977	172	.007
Number of confirming questions	.871	172	.008
Number of disconfirming questions	.871	172	.000
Average importance rating of incriminating evidence	.981	172	.018
Average importance rating of exonerating evidence	.937	172	.000

Conviction Rates

The conviction rates for Time 1 and 2 are shown in Figure 1. Binary logistic regressions were conducted to determine whether condition was a significant predictor of the decision whether or not to convict the suspect at Time 1 and Time 2. In line with the ratings for likelihood of Sabine being guilty, condition was not a significant predictor of the decision on whether or not to convict Sabine at Time 1, $\chi^2(1) = 0.00$, $p = .984$. Similarly, condition was not a significant predictor of the decision on whether or not to convict Sabine at Time 2, $\chi^2(1) = 0.14$, $p = .706$.

Figure 1

Conviction Rates for Sabine at Time 1 (After Only Reading the Case File) and Time 2 (After Receiving Answers to the Selected Questions.



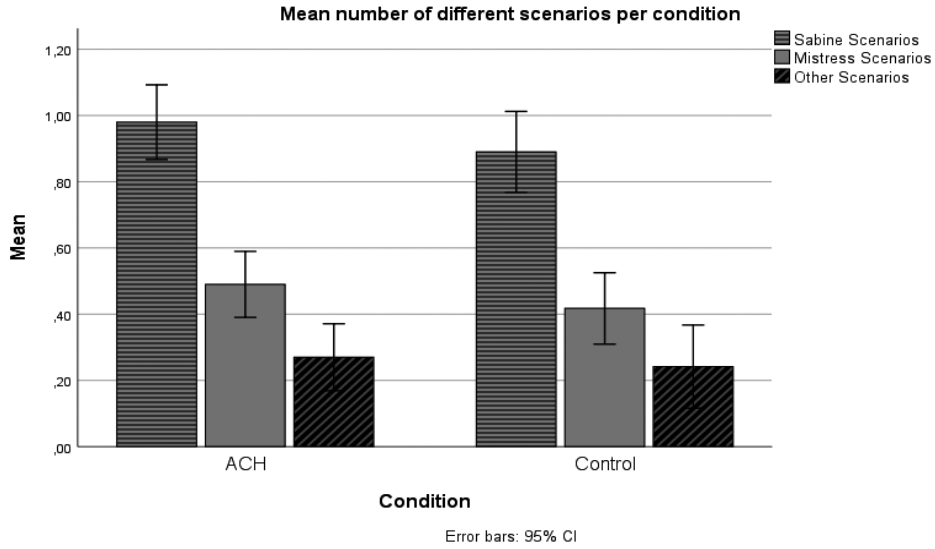
Scenarios Formulated

An exploratory analysis was also conducted for the number of scenarios formulated by participants in the ACH condition compared to the control condition. The number of scenarios that implicated Sabine or Jasper's mistress were counted. A third category was included for other scenarios, such as those

which implicated Sabine's friends who were mentioned in the case description. Overall, participants generated rather few hypotheses (see Figure 2). There was no association between condition and the formulation of at least one scenario about a perpetrator other than Sabine, $\chi^2(1) = 0.55$, $p = .458$.

Figure 2

Mean Number of Scenarios Formulated for Different Perpetrators by Participants in Both Conditions



Note. Error bars represent 95% confidence intervals.

Compliance with the ACH Instructions

As not all participants in the ACH condition followed the ACH instruction, it was also explored whether the compliance with the ACH instructions affected their responses. Those who constructed an ACH matrix were comparted to those who did not, as well as to participants in the control conditions. The results of the analyses can be found in table S4.

Table S4

Kruskal-Wallis Analyses for the Use of ACH Instructions for the Dependent Variables

Measure	*df*	X^2	*p*
Likelihood of guilt 1	2	2.04	.361
Likelihood of guilt 2	2	0.162	.922
Number of confirming questions	2	0.438	.803
Number of disconfirming questions	2	0.974	.614
Average importance rating of incriminating evidence	2	0.324	.850
Average importance rating of exonerating evidence	2	1.380	.501

Maastricht University

The House of
Legal Psychology

GÖTEBORGS UNIVERSITET